Formal grammar, usage probabilities, and English tensed auxiliary contraction

(revised June 15, 2020)

Joan Bresnan

At first sight, formal theories of grammar and usage-based linguistics appear completely opposed in their fundamental assumptions. As Diessel (2007) describes it, formal theories adhere to a rigid division between grammar and language use: grammatical structures are independent of their use, grammar is a closed and stable system, and is not affected by pragmatic and psycholinguistic principles involved in language use. Usage-based theories, in contrast, view grammatical structures as emerging from language use and constantly changing through psychological processing.

Yet there are hybrid models of the mental lexicon that combine formal representational and usage-based features, thereby accounting for properties unexplained by either component of the model alone (e.g. Pierrehumbert 2001, 2002, 2006) at the level of word phonetics. Such hybrid models may associate a set of 'labels' (for example levels of representation from formal grammar) with memory traces of language use, providing detailed probability distributions learned from experience and constantly updated through life.

The present study presents evidence for a hybrid model at the syntactic level for English tensed auxiliary contractions, using LFG with lexical sharing (Wescoat 2002, 2005) as the representational basis for the syntax and a dynamic exemplar model of the mental lexicon similar to the hybrid model proposals at the phonetic word level. However, the aim of this study is not to present a formalization of a particular hybrid model or to argue for a specific formal grammar. The aim is to show the empirical and theoretical value of combining formal and usage-based data and methods into a shared framework—a theory of lexical syntax and a dynamic usage-based lexicon that includes multi-word sequences.

Tensed auxiliary[1] contractions in English are particularly interesting because the contracting elements appear to cross the boundary between the major constituents of the sentence, namely the subject and the verb phrase, as in *You're sick*: here contracted *are* serves as the main verb of the sentence and contracts with the subject *you*. These contractions are not semantic constituents of their larger utterances. For example, the contractions *law's* and *hell's* in the sentences *my other brother in law's Arab* (authentic example from

---

[1]The term "auxiliary" includes the copula in the present study, because it shares the syntactic properties that distinguish auxiliary verbs from main verbs. These include *not* placement, *n't* contraction, and subject-verb inversion: *She is not sleeping/sleepy* vs. **She sleeps not*; *She isn't sleeping/sleepy* vs. **She sleepsn't*; and *Is she sleeping/sleepy?* vs. **Sleeps she?* See Huddleston & Pullum (2002) for discussion of the full set of distinguishing properties of auxiliary verbs.

the Buckeye corpus, Pitt *et al.* 2005) and *who the bloddy hell's knocking* (from the Canterbury Corpus, Gordon *et al.* 2004) are not compositional components in the semantics of the sentences. Nor are they syntactic constituents: witness **Who's do you think coming?* vs. *Who do you think is coming?* (Anderson 2008:174), or **It's you're that sick* vs. *It's you that are sick.*[2]

Nevertheless, tensed auxiliary contractions in some contexts show signs of being units. For example, the clitic auxiliary *'s* provides a coda of the open syllables of *law* and *hell* that select the voiced variant [z] (in contrast to *his wife's a teacher*, which selects [s]). For these reasons tensed auxiliary contraction has long been treated in formal linguistic frameworks as SIMPLE CLITICIZATION (Zwicky 1977), a phonological grouping of two adjacent non-constituent words belonging to the surface syntactic phrasings of metrical and prosodic phonology (Selkirk 1984, 1996; Inkelas & Zec 1993; Anderson 2008; Anttila 2017, Ito & Mester 2018)—purely supra-lexical phonological processes.

Yet that is far from the whole story: a number of researchers have pointed out morphophonological properties of the most common auxiliary contractions that are signs of the contracted forms being lexically stored (Kaisse 1985; A. Spencer 1991; Bybee & Scheibman 1999; Scheibman 2000; Wescoat 2005, Bybee 2006). And usage statistics show that the probability that words will be adjacent in naturally occurring speech determines their "degree of fusion" into lexical units (Bybee & Scheibman 1999, Scheibman 2000, Bybee 2002) and their likelihood of contraction (Krug 1998, Bybee 2002, Frank & Jaeger 2008, Bresnan & Spencer 2012, J. Spencer 2014, Barth & Kapatsinski 2017, Barth 2019).

What appears to be needed to explain fully the properties of tensed auxiliary contractions is a theory of their representations that simultaneously accounts for their syntactic non-constituency and adjacency constraints, their lexical morphophonology, their prosodic and metrical phonology as well as the effects of usage probability on their degree of morphophonological fusion and their likelihood of contraction. In other words, what is needed is a theory that can account for the combined findings of formal and usage-based studies of tensed verb contraction.

Unfortunately, although tensed auxiliary contraction in English is one of the empirical domains that have attracted research in both formal and usage-based

---

[2]These interrogative and clefting constructions otherwise allow larger syntactic constituents to appear with the focused phrase, as in *At what time do you think she's coming?*, *It's with Louise that she was running.*

theories of grammar, the two lines of research have proceeded mostly independently and have thus failed to provide a full answer to the deeper questions contraction poses. "Formal" research on English auxiliary contraction includes analyses in various systems of generative grammar (such as Zwicky 1970; Baker 1971; Bresnan 1971; Kaisse 1983, 1985; Zwicky & Pullum 1983; Selkirk 1984, 1996; Klavans 1985; Inkelas & Zec 1993; Wilder 1997; Sadler 1998; Barron 1998; Bender & Sag 2001; Wescoat 2002, 2005; Anderson 2008; Anttila 2017; Ito & Mester 2018). "Usage-based" research on English auxiliary contraction has included earlier work examining frequency effects on contractions (Krug 1998; Bybee & Scheibman 1999; Scheibman 2000; Bybee 2001, 2002) and more recent corpus studies of the probabilities of actual uses of contraction, employing quantitative methods such as statistical modeling of data using information-theoretic measures (for example, Barth 2011; Frank & Jaeger 2008; Spencer 2014, Barth & Kapatsinski 2017, Barth 2019). Sociolinguistic research on the topic in the Labovian tradition has generally adopted quantitative methods for modeling variation, as well as the representational basis of generative grammar, usually with the primary focus on relating the grammar of the copula to social factors (Labov 1969; McElhinny 1993; Rickford *et al.* 1991; MacKenzie 2012, 2013).

The present study of tensed auxiliary contraction proposes that the formal syntactic theory of lexical sharing in LFG, combined with a hybrid exemplar-dynamic model of the mental lexicon, can provide the necessary combined approach. Lexical sharing in LFG was originally designed to account for narrowly defined types of cases where lexical units do not match constituent structure units, such as contractions of preposition-determiner combinations (for example German *zum, am, im, ins* and French *du, au, des, aux*, discussed by Bybee 2002 and Wescoat 2007, among others), and contractions of simple clitics like English tensed auxiliary contractions (also discussed by Bybee 2002, 2010 and Wescoat 2005, among others). However, as the present study shows, lexical sharing naturally extends to the lexicalization of multi-word sequences in larger constructions.[3] While the formal analyses by themselves provide insights into the grammar of tensed auxiliary contraction, they ignore the explanatory role of usage probabilities in syntactic lexicalizations. On the other hand, usage-based linguistic studies of tensed auxiliary contraction have seldom presented fully

---

[3]Although most previous work with lexical sharing in LFG has concerned contraction, cliticization, and portmanteau-word phenomena with prepositions and determiners (e.g. Wescoat 2007, 2009; Broadwell 2008; Alsina 2010; Lowe 2016), Broadwell (2007) already extends the theory to certain multi-word expressions that form phonological words in Zapotec.

articulated proposals for their syntactic representations, leaving a wealth of systematic grammatical properties out of account. The present study therefore contributes to both formal and usage-based lines of research.

The first three sections below outline some of the main findings of usage-based linguistics on tensed auxiliary contractions and show how they are explained theoretically. The following three sections outline the main findings of formal research on tensed auxiliary contraction, and show how they are captured in the particular formal framework of lexical syntax adopted here. The next section presents a hybrid model that synthesizes the formal and usage-based findings, and the following sections present novel evidence for such a hybrid: a corpus study of *is* contraction, a formal analysis of gradient subtypes of contracting auxiliaries, and the extension of the formal grammar of auxiliary contraction to a multiword expression of classic usage-based grammar (Bybee & Scheibman 1999) that brings out surprising parallels with tensed auxiliary contraction.

A note on data sources and methods

In keeping with the goal of synthesis, the present study draws on data sources and methods from both formal and usage-based linguistics. The data consist of grammaticality judgments from the linguistic literature and the author's own speech, as well as authentic evidence from corpora. If an example is not labelled "authentic," it is constructed. The primary sources of authentic data are (1) the Buckeye Corpus (Pitt *et al.* 2005) of spoken mid-American English, and (2) the Canterbury Corpus (Gordon *et al.* 2004) of spoken New Zealand English. Quantitative datasets of variable tensed auxiliary contractions from both corpora are visualized in plots or statistically modeled. In addition, judgment data of examples are validated or corrected, where possible, using examples from Buckeye and Canterbury as well as MacKenzie's (2012) careful corpus study of auxiliary contraction in spoken English, and finally the Web, for a few rarer constructions.[4] But sometimes judgments simply represent "working evidence" to motivate a hypothesis until substantiating data can be obtained.

The Buckeye Corpus consists of one-hour interviews with each of 40 people, amounting to about 300,000 words. The speakers are Caucasian, long-time local residents of Columbus, Ohio. The language is unmonitored casual speech.

---

[4]Keller & Lapata (2003) show that the Web can be employed to obtain frequencies for unseen bigrams in a given corpus. They demonstrate a high correlation between Web frequencies and corpus frequencies, and between Web frequencies and plausibility judgments.

The data are stratified by age and gender: 20 older (defined as age 40 or more), 20 younger; 20 male, 20 female. The words and phones are aligned with sound waves, orthographically transcribed, and provided with broad phonetic labeling.

The Canterbury Corpus is a subcorpus of the Origins of New Zealand English corpora (ONZE). It consists of recorded and orthographically transcribed interviews. Speakers are born between 1930 and 1984, and interviews are added every year with the aim of filling a sample stratified by age, gender, and social class. At the time of collection of the data used in this study, the entire Canterbury Corpus consisted of 1,087,113 words.

1   Usage and phonetic reduction

A major finding of usage-based linguistics is that more probable words and multi-word expressions are phonetically more reduced and become lexically stored (Bybee 2001, 2006; Bybee & Hopper 2001; Pierrehumbert 2001, 2002, 2006; Seyfarth 2014; Sóskuthy & Hay 2017). For example, Bybee & Scheibman (1999) show that in *don't* contraction, the reduction process is most advanced with the most frequent context words and the reduced multiword forms have accrued additional pragmatic functions along with the changes in form, indicating their lexical storage as separate units from their components. These are typical effects of lexicalization: when composite items are lexically stored as wholes, they begin to acquire their own usage profiles and drift in their grammatical and semantic properties from their constituent elements.

Bybee & Scheibman (1999) collected and transcribed tokens of *don't* from about three hours and 45 minutes of "naturally occurring conversations." In Table 1, which gives excerpts from Bybee & Scheibman (1999:581–582), the words of the left and right contexts of *don't* are ordered by frequency from top to bottom. Thus pronouns, as preceding contexts of *don't*, are far more frequent than lexical NPs and among the pronouns, *I* is the most frequent. As following contexts of *don't*, the verbs *know* and *think* are the most frequent. The extent of phonetic reduction increases from left to right: the final stop deletes, the initial stop becomes a flap and then also deletes, and the vowel reduces, so that ultimately *don't* is pronounced as a nasalized schwa. As the table shows, *don't* is more highly reduced phonetically in the most frequent contexts *I*__ and __*know*, __*think*, than in all others.

According to Bybee & Scheibman (1999), these developments arise when

	[dõt, dõ]	[ɾõt, ɾõ]	[ɾə̃]	[ə̃]	Total
Preceding					
I	16	22	38	12	88
you	7	7			14
we	2	6			8
they	1	3			4
lexical NP	5				5
. . .					
Following					
know	2	8	24	5	39
think	7	6	6	1	20
have	1	7	1		9
have to	1	2	1		4
want	1	1	3		5
see	3	1			4
. . .					

Table 1: *Don't* variants by type of preceding and following item in data from Bybee & Scheibman 1999:581–582. Preceding and following contexts decrease in frequency from top to bottom; phonetic reduction increases from left to right.

frequent motor repetition in articulation becomes automatized, the automatization of pronunciation leads to blurring of word and morpheme boundaries and compression of entire multiword units; over time the result becomes a new lexically stored unit, which separately accrues its own characteristics of form and function. Lexicalization occurs because "lexical storage is highly affected by language use, such that high-frequency forms have stronger lexical representation than low-frequency forms" (Bybee & Scheibman 1999:583). As shown in Table 2 the reduced-vowel variants of *don't* in *I don't know* contrast overwhelmingly with the full-vowel variants in expressing special pragmatic functions of "indicating speaker uncertainty and mitigating polite disagreement in conversation" (Bybee & Scheibman 1999:587) in addition to the literal lexical sense.[5]

	Full vowel	Schwa
Lexical sense	7	12
Pragmatic function	1	17

Table 2: Full-vowel and reduced-vowel variants of *don't* by lexical versus pragmatic function in data from Bybee & Scheibman (1999:587).

---

[5]Applying a one-sided Fisher exact test to Table 2 to ascertain whether the odds ratio of vowel reduction co-occurring with the pragmatic function is reliably greater than 1, as predicted, yields p-value $= 0.02545$.

2 Usage and syntactic contraction

Another major finding is that the syntactic contraction, or cliticization, of word sequences is most advanced among the sequences with the highest usage probabilities. Consider tensed auxiliary contraction, which occurs when a specific set of tense-bearing auxiliary verbs, including *is, are, am, has, have, will*, and *would*, lose all but their final segments, orthographically represented as *'s, 're, 'm, 's, 've, 'll*, and *'d*, and form a unit with the immediately preceding word, called the HOST.

Although the influential early formal analysis of Labov (1969) treats the contracted verb forms as phonological reductions of the full uncontracted forms, many subsequent phonological analyses hold that synchronically, the contracted forms are are allomorphs of the full forms (Kaisse 1985; Inkelas 1991; Inkelas & Zec 1993; Anderson 2008; Mackenzie 2012, 2013). Evidence for analyzing contracting auxiliaries as morphological variants rather than phonological reductions or rapid-speech effects includes (1) the fact that there are grammatical differences between the contracted and full forms: e.g. *there's three men outside* vs. **there is three men outside* (see Dixon 1977, Nathan 1981, Sparks 1984, Kaisse 1985); (2) that phonological rules that delete the onsets and schwas of specific auxiliaries cannot be assimilated to post-lexical "rapid-speech phenomena such as deletion of flaps, coalescence of vowels etc." (Kaisse 1983:95); (3) that the phonology of specific contractions cannot be assimilated to function-word reduction in general (Kaisse 1985); and (4) that speech rate is not predictive of auxiliary contraction in spoken corpus data (Frank & Jaeger 2008). It is also worth noting that auxiliary contraction cannot simply be assimilated to casual speech (McElhinny 1993:376): in style-shifting among white speakers, *is* contraction occurred 79% of the time in casual speech (in group interviews) and 87% of the time in careful speech (in single interviews) (Labov 1969:730–731).

A usage-based corpus study of tensed auxiliary contraction in "spoken mainstream British English" by Krug (1998) finds that the contraction of tensed auxiliary verbs (e.g. *I've, he's, we'll*) varies directly with the BIGRAM PROBABILITY ("string frequency") of the subject and the auxiliary. Even where the preceding phonological contexts are similar—open monosyllables ending in tensed vowels in *I've, you've, we've, they've, who've*—the bigram probability directly correlates with the proportions of contractions.[6]

---

[6]In a corpus study of contractions in the Switchboard corpus (Godfrey & Holliman 1997), MacKenzie (2012:130,149–155) finds that the frequency effect on contraction "does hold for the extreme ends of the frequency scale (i.e., the most and least frequent host/auxiliary

Recent work on several other varieties of spoken English has confirmed the basic finding that probabilistic measures derived from frequencies of use of hosts and auxiliaries correlate with the likelihood of contraction (Frank & Jaeger 2008, Barth 2011, Bresnan & Spencer 2012, J. Spencer 2014, Barth & Kapatsinski 2017, Barth 2019). These works employ counts of the frequency of use of host-auxiliary sequences to estimate their probabilities, from which they calculate *transition probabilities, conditional probabilities, informativeness*, and related measures.

The measure adopted in the present study is the negative logarithm of the conditional probability of the host given the auxiliary. The conditional probability of $word_1$ appearing before $word_2$ (1a) in some language can be estimated from a particular corpus by the calculation shown in (1b). The inverse of the conditional probability is its reciprocal, which grows smaller as the probability grows larger, and approaches zero as the probability approaches one; inversely, very low probabilities yield extremely high values in the reciprocal. The logarithm of this inverse, which compresses extreme values, yields (1c), here termed the INFORMATIVENESS.[7]

(1) (a) Conditional probability: (b)   Estimated:
$$P(word_1|word_2)$$ $$\frac{count(word_1 word_2)}{count(word_2)}$$

(c) Informativeness: $$log\frac{1}{P(word_1|word_2)} = -logP(word_1|word_2)$$

Why choose a measure of usage probability of the host given the *following* word, and not the *preceding* word? One answer is that probability conditioned

combinations do contract at a high and a low rate, respectively), but that the string frequency/contraction connection does not hold to any degree of granularity in the middle," concluding that "the attested pronoun-specific effects on short allomorph selection cannot be explained by string frequency alone." Her results are based on (estimated) raw string frequencies, as are the findings of Krug (1998). Research discussed below supports the effects of conditional probabilities of contraction with a host in the contexts of specific auxiliaries.

[7]In information theory this quantity is known as the SURPRISAL (Shannon 1948). The IN-FORMATION CONTENT of a word $word_i$ in information theory (Shannon 1948) is the weighted average of its surprisal $-log_2 P(word_i|context_j)$ over all contexts j; however, in cases of a single specific linguistic context of interest (such as a simple clitic), averaging makes no difference to the value. The present study uses the term "informativeness" in this case. The "informativeness" of a host in the context of an auxiliary is also proportional to its joint probability with the auxiliary—in other words, their probability of occurring together. This latter is Bybee's (2001, 2010) measure of usage probability.

on the next word can be viewed as measuring the LEXICAL ACCESSIBILITY of $word_1$ in the context of the speaker's planned next word, $word_2$: the ratio measures $word_1$'s share of all tokens that precede $word_2$; it thus corresponds to $word_1$'s relative availability or activation in that context. The probability of the auxiliary given the *preceding* word $P(word_2|word_1)$ would presumably be more helpful to the listener, who does not have access to the speaker's planned next word.[8] Another answer is that conditional probability derived from the following context is often a better predictor than that derived from the preceding context in speech production processing data (Ernestus 2014).

Figure 1 plots the relation between informativeness and contraction of present tense *have* and *be* with pronominal hosts from the Buckeye Corpus.[9] Figure 1 clearly shows a strong inverse relation between the log likelihood of contraction and informativeness of the pronoun hosts before the verb forms: the first person singular pronoun *I* has the least informativeness before the first person singular verb form *am*, and that sequence has the highest log likelihood of contraction. As informativeness increases from left to right, the log likelihood of contraction shows a steady decrease for present-tense forms of both *be* and *have*.

3   The mental lexicon

What explains the close relation between usage probability and contraction? Krug (1998:305) hypothesizes that the word or sequence of words in subject-auxiliary contractions is stored in the mental lexicon, which responds dynamically to usage probabilities as proposed by Bybee (1985:117):

> "Each time a word is heard and produced it leaves a slight trace
> on the [mental] lexicon, it increases its lexical strength."

Pierrehumbert's (2001, 2002, 2006) exemplar-dynamics model fleshes out

---

[8]Several recent studies provide evidence from both corpora and experiments in favor of an accessibility-based model of speech over a model based on uniform information density which could be interpreted as favoring the hearer (Zhan and Levy 2018, 2019).

[9]The data points in Figure 1 represent 7614 present tense *be* forms and 805 present tense *have* forms collected ANON respectively by J. Spencer (2014) and the author from the Buckeye Corpus by the orthographic transcriptions *have/ 've, has/'s, am/'m, are/'re,* and *is/'s*. Instances in which the grammatical context did not permit contraction were excluded following MacKenzie (2012). The remaining instances were checked against their phonetic transcriptions to ensure that orthographically contracted auxiliaries corresponded to phonetically asyllabic forms. Informativeness was calculated as in (1).
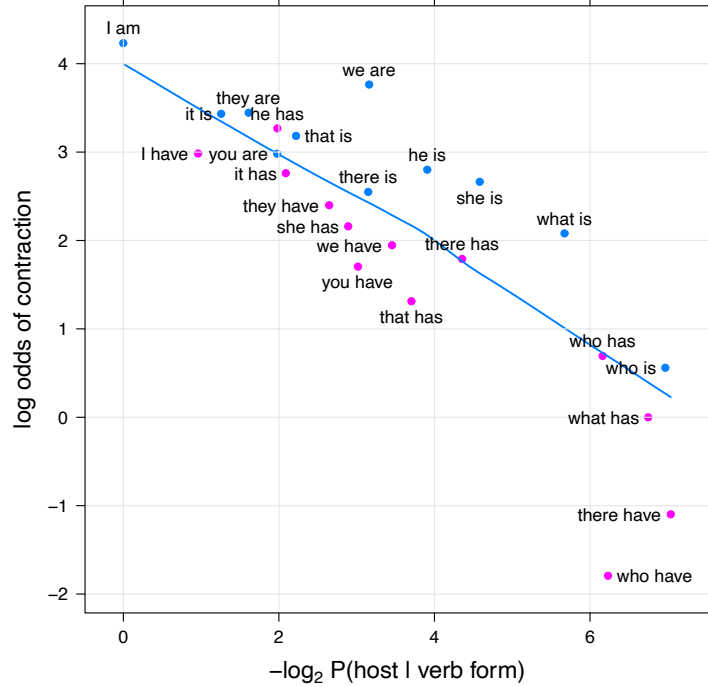
Figure 1: Relation between the log odds of contraction and the negative log conditional probability ('informativeness') of pronoun hosts in the context of verb forms in the Buckeye corpus. The *have* and *be* datasets are respectively plotted with magenta and cyan dots, with a loess smoother showing the trend in the combined data.

this concept of the mental lexicon: it consists essentially of a map of the perceptual space and a set of labels, or structural descriptions, over this map. Long-term memory traces are located in the perceptual space and clustered by similarity. Each exemplar has an associated strength or resting activation; exemplars of frequent recent experiences have higher resting activation levels than those of infrequent and temporally remote experiences.

In this model speech perception involves the labeling of new instances based on their similarity to existing instances stored in memory, and speech production involves randomly selecting a target exemplar from the same space of stored memory instances; the production of that target is then added to the store of exemplars. This is the PERCEPTION-PRODUCTION LOOP, which dynamically

affects language change by amplifying slight biases or changes over many iterations. For example, a slight but constant production bias toward lenition in each utterance can result in gradual sound changes in which more frequent words show a higher rate of change than less frequent (Phillips 1984, Bybee 2000): more frequently uttered words refresh their stores of lenited exemplars while the exemplars of less frequently uttered words are selected less often as targets of production because of the greater impact of memory decay (Pierrehumbert 2001).[10,11] Applied to multiword sequences, the model can simulate the relation between usage and phonetic reduction (Section 1). Applied to host + tensed auxiliary sequences, the model can also simulate the relation between usage probabilities and contraction (Section 2).

Figure 2 provides a simplified visualization of tensed auxiliary contractions in this model. The LABELS *you, you're,* and *are* with their varying pronunciations stand for (partial) 'lexical entries' in traditional linguistic terminology and correspond to structural descriptions at several levels, not shown (see Wright *et al.* 2005, German *et al.* 2006). Each entry maps onto a matching set of remembered instances of its utterance—the MEMORY TRACES (or EXEMPLARS). The visualization is simplified to show only varying pronunciations of remembered instances; it omits links to further grammatical, pragmatic, semantic, and social information. Fresh experiences and memory decay lead to continual updating of the entries in the mental lexicon.

The mental lexicon stores both words and multi-word fragments (Bybee & Scheibman 1999, Bybee 2010). Among the multi-word fragments would be *you are*, the uncontracted sequence of function words that is functionally equivalent to *you're* in grammatical structure (although they may of course differ in other properties such as prosody, discourse context, and speaker style). Instances of both would have a common label at some level of grammatical labeling. In this way the mental lexicon would implicitly encode bigram probabilities and informativeness as activation levels of the various words and multi-word

---

[10]The theoretical types of frequency effects generated by the model depend on the parameter range for memory decay and are broader than discussed here. More recent work has developed dynamic exemplar models further to incorporate the perceptual biases of the listener (Todd 2019, Todd *et al.* 2019).

[11]The relation between production and perception assumed here is obviously simplified. Further, there is evidence that word frequency effects vary with the production or perception task (Harmon & Kapatsinski 2017) and that 'word prevalence'—how many different people use a word—may be a better estimate of frequency effects on lexical decision times (Brysbaert *et al.* 2016).
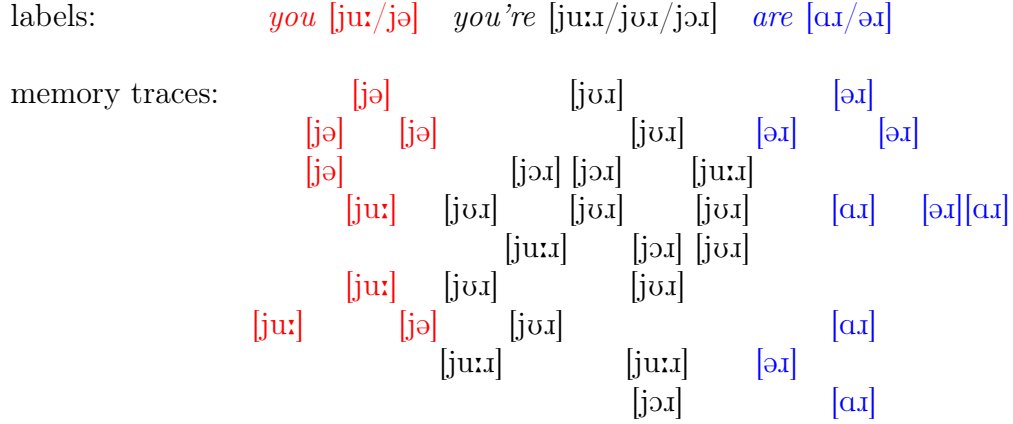
labels: *you* [juː/jə]   *you're* [juːɹ/jʊɹ/jɔɹ]   *are* [ɑɹ/əɹ]

memory traces: [jə]   [jʊɹ]   [əɹ]
                   [jə]   [jə]   [jʊɹ]   [əɹ]   [əɹ]
                   [jə]   [jɔɹ] [ɹɛ]   [juːɹ]
                        [juː]   [jʊɹ]   [jʊɹ]   [jʊɹ]   [ɑɹ]   [əɹ][ɑɹ]
                                        [juːɹ]   [jɔɹ] [jʊɹ]
                        [juː]   [jʊɹ]   [jʊɹ]
              [juː]   [jə]   [jʊɹ]   [ɑɹ]
                        [juːɹ]   [juːɹ]   [əɹ]
                        [jɔɹ]   [ɑɹ]

Figure 2: Visualization of tensed auxiliary contractions in an exemplar-dynamic model of the mental lexicon (Pierrehumbert 2001, 2002, 2006), which includes memory traces of speech events of varying activation levels (a function of the density of exemplars under each label as well as their recency).

fragments that are stored there.

In contraction, the short (asyllabic) allomorphs of the auxiliaries are phonologically incorporated into the final syllable of the host.[12] Assuming a production bias favoring the short allomorph parallel to the production bias favoring lenition, the crucial connection between high-probability (low-informativeness) host-auxiliary bigrams and higher incidences of contraction in speech production is then straightforward: the more frequently uttered bigrams refresh their stores of contracted exemplars while those of less frequently uttered bigrams are more temporally remote, lower in activation, and less likely to be selected as targets of production.

If highly probable contractions are lexically stored with phonetic detail, they should accumulate allophonic reductions as part of their long-term representations (Section 1). As Bybee (2006:723) puts it, "Frequent phrases such as *I don't know, I don't think*, and many others show phonological reduction in excess of that which could be attributed to on-line processes, such as that evident in other tokens of *don't*, as in *I don't inhale*, indicating that such reduction has accumulated in representation." There is evidence that fits this expectation.

Wescoat (2005:471) gives various examples of "morphophonological idiosyn-

---

[12]Bybee (2002:124–5) demonstrates that spoken usage frequencies favor encliticization over procliticization of the asyllabic auxiliary.

cracies" among tensed auxiliary contractions, shown in Table 3. One of them is that "*I* [aɪ] may be pronounced [ɑ], but only in association with *'ll* (*will*), yielding [ɑl]; moreover *you* may become [jɔ], but only when followed by *'re* (*are*), resulting in *you're* [jɔɹ]."[13] Thus the reduced pronunciations are specific to individual pronoun-auxiliary sequences. He emphasizes that these pronunciations are not fast-speech phenomena: *I'll* [ɑl] and *you're* [jɔɹ] "may be heavily stressed and elongated." In other words, their pronunciations are not merely on-line contextual adjustments to the phonology of rapid connected speech.

I'll	[aɪl/**ɑl**]	*I'm*	[aɪm/*ɑm]	*I've*	[aɪv/*ɑv]
you'll	[juːl/*jɔl]	*you're*	[juːɹ/**jɔɹ**]	*you've*	[juːv/*jɔv]

Table 3: Contrasting contraction-specific pronunciations from Wescoat (2005)

Diachronically, these pronunciations could theoretically derive from such on-line contextual adjustments of frequently repeated sequences (for example, the velarization or darkening of /l/ in *will* and the laxing of immediately preceding unstressed vowels, yielding *we'll* [ˈwiː.əl, wɪl]). But the retention of the reduced pronunciations of specific words even in slow or emphatic speech shows that synchronically, their distribution does not match that of on-line contextual adjustments to the phonology of rapid connected speech. It rather supports lexical representation of the reduced variants. The simplest account is that synchronically they are lexically stored allomorphs of the host + auxiliary.

Along the same lines, Piantadosi *et al.* (2011) show from a cross-language corpus study that information content is an important predictor of orthographic word length (more so than raw frequency), across lexicons from a variety of languages:

> One likely mechanism for how the lexicon comes to reflect predictability is that information content is known to influence the amount of time speakers take to pronounce a word: words and phones are given shorter pronunciations in contexts in which they are highly predictable or convey less information [references omitted]. If these production patterns are lexicalized, word length will come to depend on average informativeness.

---

[13]Here he describes his own speech, but notes that Sweet (1890:25) also reports this pronunciation of *you're*, and it is shared by the present author as well.

The Bybee-Pierrehumbert theory of the mental lexicon provides an explicit model of the lexicalization of production patterns in which more probable (less informative) words become reduced (shorter).[14,15]

4 The grammatical contexts of contraction

Studies of the grammatical contexts that permit or prohibit contraction—particularly their syntax and prosody—have provided the main findings of research on the topic in formal linguistics. Yet despite its explanatory depth, usage-based linguistics has not provided a detailed understanding of these contexts.[16] The following three subsections summarize those findings most relevant to the present study. For these it is useful to distinguish between unstressed syllabic and asyllabic forms of the tensed auxiliaries as in Table 4 adapted from Inkelas & Zec (1993) and Wescoat (2005), who follows Sweet (1890:14–16).[17]

Metrical dependence on the right context

The asyllabic forms of contracted tensed auxiliaries share metrical constraints on their right contexts with the unstressed syllabic forms of the same auxiliaries. This relation is what Selkirk (1984:405) describes as "the central generalization" about auxiliary contraction: "only auxiliaries that would be realized as stressless in their surface context may appear in contracted form." It is also the

---

[14]Seyfarth (2014) discusses this and possible alternative models of the effects of informativeness, or average contextual predictability, on lexicalization of words' durations. All of the alternatives he discusses but one assume with Bybee and Pierrehumbert that both reduced forms and their probabilities of use are lexically stored; hence, all of these alternatives are broadly consistent with the hybrid formal/usage-based approach described here, and may be regarded as variant models of the fundamental usage-based insight connecting lexicalization with probability and reduction. One alternative Seyfarth proposes assumes that only word-specific probabilities and not reduced forms themselves are stored, but that proposal would not very naturally account for the accrual of lexically specific phonetic, semantic and pragmatic accruals of the kind found by Bybee & Scheibman (1999) (see Tables 1 and 2).

[15]See Bybee & McClelland (2005) for discussion of a distributed connectionist alternative model and Ambridge (2019) for a broad discussion of exemplar theories and alternative models in an acquisition context.

[16]—although Barth and colleagues analyze contraction by broad construction type such as copula, future, and progressive (Barth 2011, Barth & Kapatsinski 2017, Barth 2019).

[17]In the present study [ə] represents the stressless mid-central vowel and [ɨ] represents a slightly higher unstressed vowel.

	full ("strong")	unstressed syllabic ("weak")	asyllabic ("enclitic")
are	[ɑɹ]	[əɹ/ɹ̩]	[ɹ]
am	[æm]	[əm]	[m]
had	[hæd]	[(h)əd]	[d]
have	[hæv]	[(h)əv]	[v]
has	[hæz]	[(h)əz]	[z/s]
is	[ɪz]	[ɨz/əz]	[z/s]
will	[wɪl]	[əl/l̩]	[l]
would	[wʊd]	[(w)əd]	[d]

Table 4: Strong, weak, and enclitic forms of the tensed auxiliaries

core generalization of Labov's (1969) analysis, which phonologically derives the asyllabic forms from the syllabic.

The right context of both syllabic and asyllabic reduced auxiliaries requires that the auxiliary be followed by a stressed word, as (2a,b) illustrate.[18]

(2) a. *They are/*'re __.* [ðeɪ ɑɹ/*ˈðeɪ.əɹ/*ðɛɹ]

b. *They are/'re here.* [ðeɪ ɑɹ/ˈðeɪ.əɹ/ðɛɹ]

The stressed word need not be adjacent to the auxiliary. In line with Labov's (1969) observations, *is* reduces and contracts before the nonadjacent stressed verb *doing* in (3a), but not before unstressed *it* alone:[19,20]

(3) a. *That bird, what's it doing __?* [wʌts ɨt ˈduɪŋ]/[ˈwʌt.əz/ɨz ɨt ˈduɪŋ]

b. **That bird, what's it __?* *[wʌts ɨt]/*[ˈwʌt.əz/ɨz ɨt]
(cf. ..., *what* IS *it?/what's* IT*?*)

Stressed constituents falling outside of the complement phrase of the auxiliaries do not support contraction (Labov 1969). In (4), for example, Inkelas & Zec (1993:234) analyze the temporal adverbs as outside the complement phrase of the reduced or contracted *is*:

---

[18]Following Wescoat (2005), the dot '.' marks a syllable boundary.

[19]MacKenzie (2012:79–82) cites spoken corpus data showing the same effect with several unstressed pronouns.

[20]This generalization applies to unstressed referential pronouns; contraction before unstressed *it* does occur phrase finally in some fixed expressions, such as *whosie-whatsit*, an Australian slang term for someone or something whose name has been temporarily forgotten, and *howsit, howzit*, a New Zealand slang greeting.

(4) *I don't know where the party is* [ɪz/*ɨz/*z] __ *tonight.*

Since the publication of a squib that excited decades of work on the syntax of auxiliary contraction (King 1970), many linguists have continued to judge contraction to be blocked before pre-focus gaps and ellipsis medially within the complement phrase, as in the "comparative subdeletion" (5) and "pseudo-gapping" (6) examples from Inkelas & Zec (1993). In these examples, the words in small caps are uttered with pitch accents, indicating parallel foci contrasting both subject and object.

(5) KAREN *is a better* DETECTIVE *than* KEN *is/*'s* __ *an* ARCHEOLOGIST.

(6) JOHN*'s playing* ROULETTE*, and* MARY *is/*'s* __ BLACKJACK.

To account for the apparent ungrammaticality of contraction in such examples, many analyses have hypothesized that contraction is blocked within the verb phrase before a medial syntactic gap or ellipsis for various reasons (e.g. Bresnan 1971; Kaisse 1983, 1985; Inkelas & Zec 1993; Anderson 2008).[21] But Selkirk (1984) suggests that such phrase-medial sites of contraction actually allow variable contraction in usage, citing (7).

(7) *Looks as good as it's* __ *fun to play* (Selkirk 1984:443, n.25)

And there are authentic examples on the Web in support of this suggestion:

(8) a. *"But I know **he's a better runner than he's a biker**," Lopez said.*

   b. *. . . the spherical earth . . . shows Australia as being **4 times as as long as it's wide**, . . .*

   c. *I still think **he's a better drummer than he's a singer** but don't tell him that.*

   d. *If **it's longer than it's wide**, then it's phallic. If **it's not longer than it's wide**, then you put it on its side. Now **it's longer than it's wide**, and it's phallic!*

---

[21]The gap in (5) is supposed to correspond to an implicit degree modifier such as *(how good)* or *(that good) an archeologist* (Bresnan 1973).

16

Interestingly, these examples differ from those judged ungrammatical ((5) and (6)) in that instead of contrasting two pairs of foci—the subject and object complement of each clause—they contrast only one—the complement; the subject of each second clause is anaphoric and not in contrast. Yet the examples do not differ in the relevant pre-focus syntactic structure, so that cannot be what prevents contraction.

To account for the variability of contraction before the medial sites of deletion and ellipsis, Selkirk (1984:374ff) makes the plausible proposal that retention of the unreduced auxiliary pre-focus is one of a suite of stylistic metrical options that speakers may use to highlight prosodic and structural parallelism in constructions like those in (5) and (6) above. That is the position adopted in the present study.

Inkelas & Zec (1993) also point to examples like (9a,b), where contraction can occur directly before a gap (a), provided that a stressed complement word follows (contrast (b)):

(9) a. *I don't know how much there is/'s ___ left in the tank.*

b. **I don't know how much there is/*'s ___.*

Similar examples occur in the Web showing contractions adjacent to the extraction sites:

(10) a. *Hi, Soon going to London, and I've got an Oystercard from last time. Is there any possibility to see **how much there's left on it** and/or top up online?*

b. *So many have chimed in on Lin at this point that we're not even sure **how much there's left to say**.*

c. *No clue **what there's ___ going on.***

d. *. . . they were probably aware of **what there's ___ going on with her with the fandom**.*

The main finding important to the present study is that the unstressed tensed auxiliary forms (both syllabic and asyllabic) are metrically dependent on their complement constituents to the right.

Note that there are enclitics and weak function words that are not rightward metrically dependent and hence can occur phrase-finally. Compare the tensed auxiliary in (11a) with a possessive enclitic in (11b), a weak object enclitic in (11c), and an untensed auxiliary enclitic in (11d).

(11) a. *Who's very polite? *Tom's. (=Tom is)*

b. *Whose child is very polite? Tom's. (= Tom's child)*

c. *Kill 'em.* [ˈkɪl.m̩]

d. *I might've.* [ˈmaɪt.əv]

Enclisis with the left context

While sharing their metrical dependence on a stressed complement in the right context, the asyllabic and unstressed syllabic auxiliaries diverge with respect to the left context. Specifically, the asyllabic tensed auxiliaries form a PHONO-LOGICAL WORD with their hosts to the left, unlike their syllabic counterparts. The phonological wordhood of tensed auxiliary contractions is supported by (i) the progressive voicing assimilation of *'s* with the final segment of the host, together with (ii) the absence of pausing and interruptions between the host and the contracted auxiliary.

 Examples (12) and (13) illustrate the phenomenon of voicing assimilation (i). The choice of the specific pronunciation of *'s* depends on the phonology of the host x. Morphophonologically, *'s* contractions undergo word-internal rules of voicing assimilation or epenthesis—or perhaps more accurately, phonologically conditioned allomorph selection among the variants [z/s/ɨz]—parallel to plural and tense inflections:

(12) a. **plurals:** *peats* ([s]), *reds* ([z]), *losses* ([əz])

b. **present tense:** *bleats* ([s]), *shreds* ([z]), *tosses* ([əz])

c. **'s contractions:** *Pete's* ([s]) *here, Fred's*([z]) *here, Ross's* ([ɨz]) here

The contrast with arbitrary adjacent syntactic words shows that the voicing assimilation and epenthesis are word-internal effects specific to contractions with the auxiliary *'s*:

(13) *Pete sang* ([s]) *and Fred sang* ([s/*z])
 Fred zigged ([z]) *and Pete zagged* ([z/*s])
 Ross zig-zagged ([z/*əz])

 As for (ii) above, the authentic examples in (14a–c) are provided by MacKenzie (2012:76–79) to illustrate that contraction of the auxiliary is not found in "pseudo-clefts", "*th*-clefts", and "*all*-clefts":

(14) a. *What I'm talking about is* [ɪz] *the people over here and over here and across the street.*

    b. *Well, the problem is* [ɪz], *that most of the record players now will not play them.*

    c. *All I know is* [ɪz] *I didn't vote for him.*

(14a–c) can be thought of as "specification constructions" in that the post-auxiliary constituent specifies the meaning of the subject, as though preceded by a colon. The specification appears to form a focused phrase, which can be set off by a pause.

Inkelas & Zec (1993:243,245) propose a phonological explanation that could apply to such specification constructions as well as other authentic pre-auxiliary contexts that block extraction from MacKenzie (2012)[22] and constructed examples from Kaisse (1979, 1983, 1985). Inkelas & Zec (1993) assume that English auxiliary clitics form a phonological word ω with a phonological word to their left. Then they assume with Sells (1993) that certain focused syntactic constituents are set off by a phonological or intonation phrase boundary which prevents auxiliary encliticization.[23] How this proposal would apply to (14) is illustrated in (15).

(15) *{ *What I'm talking* (**about** } { **'s**)$_\omega$ *the people over here ...* }

If contractions are enclitics on their hosts to the left, they cannot be interrupted by pauses or by the prosodic boundaries of certain focused or dislocated syntactic phrases. In contrast, the tensed weak SYLLABIC auxiliaries are not enclitics but are phonologically dependent on their RIGHTWARD phrasal context only (Inkelas 1991; Inkelas & Zec 1993; Selkirk 1984, 1996). Hence pauses and strong prosodic boundaries can separate them from the preceding word:

(16) { *What I'm talking about* } {*is* [ɪz] *the people over here and over here and across the street* }

---

[22]These would include parentheticals, adverbs, and preposed prepositional phrases (locative inversions).

[23]They assume that the strong prosodic boundary is obligatory, but for other speakers it appears to be an optional variant. Individual or stylistic variability in the strength of prosodic boundaries would explain contrasting grammaticality judgments of constructions like *Speaking tonight is/*'s our star reporter* (cf. Inkelas & Zec 1993, p. 245 and Anderson 2005, p. 71).

(17) { *They—bicycle cranks, I mean—are* [əɹ] *expensive* }

In sum, tensed ASYLLABIC contractions are simultaneously prosodified both to the left, as part of a phonological word with the host, and to the right, like the tensed weak syllabic auxiliaries, in being metrically dependent on their complement phrases. In other words, while both the syllabic and asyllabic forms are 'proclitic' in a purely metrical sense (cf. Bresnan 1971, Wilder 1997), only the asyllabic form also encliticizes to its preceding host.

Restrictive and nonrestrictive auxiliaries

While all asyllabic tensed auxiliaries share the properties of metrical dependence on their rightward complements and enclisis on their leftward hosts, further grammatical differences divide them into subtypes that Wescoat (2005) terms RESTRICTIVE and NONRESTRICTIVE. The following classification of asyllabic forms of the tensed auxiliaries is adapted from Wescoat (2005).[24]

(18)		restrictive	nonrestrictive
	are	're	
	am	'm	
	had		'd
	have	've	
	has		's
	is		's
	will	'll	
	would		'd

According to Wescoat (2005), the restrictive asyllabic auxiliaries contract only with pronoun and *wh-* pro-form hosts, while other asyllabic auxiliaries are not restricted in this way. His examples (19a–e) show restrictive asyllabic auxiliaries with pronoun and *wh-* pro-form hosts:

(19) a. *I'll help.* [aɪl]

b. *We're a big group.* [wiːɹ]

---

[24]Wescoat follows Spencer's (1991: 383) classification of *'d* as restrictive, but notes that it contracts with non-pronoun hosts in Zwicky's (1970) and his own speech, indicating a possible dialectal difference with Spencer's British English variety. The nonrestrictive classification of *'d* is adopted here, because it accords with the author's variety of American English.

c. *They've gone.* [ðeiv]

d. *I'm happy.* [aɪm]

e. *How've you been?* [haʊv]

Wescoat (2005) constructs minimal pairs to (19) using monosyllabic non-pronoun hosts which he judges ungrammatical when pronounced with asyllabic contraction. Examples include *Ai'll help* [ai.l̩/*aɪl], *The Cree're a big group.* [kɹiː.ɹ̩/*kɹiːɹ], *The Au've been polled* [aʊ.əv/*aʊv], and *So am I* [soʊ.m̩/*soʊm].[25,26]

The nonrestrictive asyllabic auxiliaries corresponding to *is, has, had* and *would* can all contract with both pronoun and non-pronoun hosts in some varieties of American English, as the following examples slightly adapted from Wescoat (2005), illustrate.

(20) a. *It's gone/going.* [ɪts]

b. *Pat's gone/going.* [pæts]

(21) a. *She'd seen it.* [ʃhiːd]

b. *Lee'd seen it.* [liːd]

(22) a. *I'd have seen it.* [aɪd]

b. *Bligh'd have seen it.* [blaɪd]

There is a further syntactic difference between restrictive and nonrestrictive auxiliaries, illustrated by examples (23a–c) from Wescoat (2005): the hosts of the former cannot be conjuncts or occur embedded within a larger subject phrase.

(23) a. [*She and **I**] 'll help.* [aɪː.l̩/*ail/*ɑl]

b. [*The people beside **you**] 're going.* [juː.ɹ̩/*juːɹ/*jʊɹ/*jɔɹ]

___

[25]According to Wescoat (2005), *Ai* is a Japanese given name and *the Au* refers to speakers of a language of Papua New Guinea (see Simons & Fennig 2018).

[26]Wescoat's theory is compatible with a range of varying judgments, because it depends on lexical features of the host. For example if Wescoat (2005) had categorized *so* as a pro-form rather than an Adverb, it could allow contraction with restrictive asyllabic auxiliaries, and indeed a reviewer made this judgment. Further variations are discussed in Section 9.

c.   [*The people who helped* **you**]*'re kind.* [juː.ɹ̩/*juːɹ/*jʊɹ/*jɔɹ]

In contrast, the following authentic spoken examples (24a,b) from the Canterbury Corpus (Gordon *et al.* 2004) and (24c,d) from the Buckeye Corpus (Pitt *et al.* 2005) illustrate that nonrestrictive *'s* can contract with noun hosts that are dependents of the subject of the auxiliary and conjuncts:

(24) a. [*the computer science department at* **Canterbury**]*'s* [z] *really lousy*

      b. [*anything to do with* **money**]*'s* [z] *good*

      c. [*everybody in my* **family**]*'s* [z] *mechanically inclined*

      d. [*August September and* **October**]*'s* [z] *just gorgeous*

Although authentic examples are rarer, other nonrestrictive auxiliaries may not contract as freely as *'s*. (25) shows the single instance of *'d* contracted with a non-pronoun host in 2890 occurrences of contracted and uncontracted *did, had, would* in the Buckeye Corpus:[27]

(25)   *Wexner Center'd* [ˈsɛntɚd] *<SIL> be one of my primary ones.*

Judgments of constructed data are uncertain, but (26a–b) suggest that both *would* and *had* can contract with a host embedded within a larger subject phrase, at least in the author's speech:

(26) a. [*Everybody in my* **family**]*'d agree.* [ˈfæm(ə)lid] *d < would*

      b. [*Everybody in my* **family**]*'d agreed to it.* [ˈfæm(ə)lid] *d < had*

In sum, beyond their shared prosodic and metrical properties, the contracting auxiliaries appear to differ in their selectivity for the host words and their restrictiveness toward host phrases. The restrictive auxiliaries require that the host be a subject pronoun or *wh-* pro-form not embedded within a larger subject phrase. The nonrestrictive lack both of these requirements and very freely encliticize to their adjacent hosts, even nonsubjects.

---

[27]In the Buckeye Corpus transcriptions (Kiesling *et al.* 2006:19), <SIL> labels silent regions between words in separate sections of running speech, or during a speech disfluency of some kind, such as a restart or hesitation. Within a section of running speech, <SIL> labels silence of 50ms or more between words.

5 Lexical sharing

Tensed auxiliary contractions, with their morphophonological evidence for host + auxiliary allomorphy, lexical selection of the host, and varying restrictions on host phrases, are problematic for the traditional view of contraction as prosodic enclisis, as Wescoat (2005) argues. When viewed as purely phonological phrasings of two adjacent non-constituent words in the surface syntax, they are not fully accounted for by theories of metrical and prosodic phonology (e.g. Selkirk 1984,1996; Inkelas & Zec 1993; Anderson 2008; Anttila 2017, Ito & Mester 2018). But Wescoat also argues against lexicalist counteranalyses which propose that the pronoun + restrictive auxiliary contractions have been morphologized into affixed words, for example Sadler's (1998) LFG analysis and Bender & Sag's (2001) HPSG analysis, drawing on Spencer (1991:383). The essential problem is that the contractions appear to be morpholexical units but do not also behave like syntactic and semantic units. They cannot be conjoined together, and they permit coordination of the auxiliaries together with their verb phrases, as the examples in (27) illustrate:

(27) a. *[***They're*** *and* ***you're***] *going.*

b. ***I***[***'m*** *looking forward to seeing you*] *and* [*will be there on Sunday*]

c. ***You***[***'ll*** *do what I say*] *or* [*will suffer the consequences*]

The theory of lexical sharing in LFG (Wescoat 2002, 2005) provides a formal analysis of tensed auxiliary contractions in English that solves these problems, turns out to be highly compatible with usage-based findings for these phenomena, and is also broadly extendable. In this theory, morphological and phonological units do not have to be associated with just one terminal category node in the syntactic structure, but can be shared between two linearly adjacent terminal category nodes. Figure 3 provides an illustration of the idea.[28]

In Figure 3 the arrows pointing to words represent a formal mapping from syntactic constituent structures (c-structures in LFG) to the lexical items that

---

[28]The particular category names are not important; here Wescoat follows the c-structure theory outlined by Bresnan (2001) (also Bresnan *et al.* 2015), but any appropriate category labels will do. The intuition behind D and I is that these are function word categories corresponding to bleached nominals and verbs (Bresnan 2001). In early work, Postal (1966) observes that pronouns behave like determiners in English phrases like *we men, you guys*, and German anaphoric uses of *die, der* also support the D analysis of pronouns more generally.
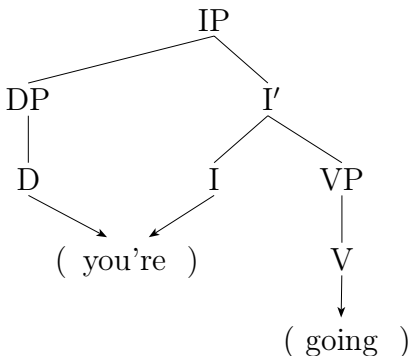
Figure 3: Relation between c-structure and tensed auxiliary contraction under lexical sharing (Wescoat 2005)

instantiate them—their LEXICAL EXPONENTS, in Wescoat's (2005) terms. As usual in LFG the c-structure represents the "surface" syntactic groupings of words, while the "deeper" relations and dependencies are provided in a parallel functional structure (f-structure) that bears many similarities to dependency grammar graphs (Mel'čuk 1988, Bresnan 2016). The surface words themselves provide most of the global functional information in the form of relational features that give rise to descriptions of the f-structure context of the word. Language-particular c-structure configurations provide what structural information about linguistic functions there may be in a given language, which in the case of configurational languages like English is fairly redundant (Bresnan *et al.* 2015).

Wescoat (2005) initially applies the lexical sharing analysis to the restrictive contractions: "The nonsyllabic contractions of *am, are, have,* and *will* (and for some speakers, *had* and *would*) are attached to pronouns and *wh*-words IN THE LEXICON" (Wescoat 2005:482). In the lexicon these restrictive contractions are associated with adjacent syntactic terminal categories and may specify item-specific phonology and functional restrictions, as illustrated in (28).[29] In (28)

---

[29]The 'down' arrows in (28) are standard LFG metavariables which give rise to functional structures when instantiated in the syntactic context of a particular sentence, phrase, or fragment of language. The double down arrow ⇓ is a special metavariable defined by Wescoat (2005) to refer to the f-structure of the lexical exponent of a category. In the case of a contraction like *you're* in (28), which is the lexical exponent of two adjacent categories, the double down arrow allows properties of the f-structure of the contraction as a whole to be specified in addition to the standard properties of the f-structures of its atomic D and I

the lexical entry for *you're* specifies the pronunciations indicated and shows that the contraction is lexically shared by the sequence of adjacent categories D and I.

(28) Lexical entries for the structure in Figure 3:

you're [juːɹ/jʊɹ/jɔɹ] ←

D	I
$(\downarrow \text{PRED}) = \text{`PRO'}$	$(\downarrow \text{TENSE}) = \text{PRES}$
$(\downarrow \text{PERS}) = 2$	$(\downarrow \text{SUBJ NUM}) = \text{PL}$
$\Downarrow = \downarrow$	$(\downarrow \text{SUBJ}) =_c \Downarrow$

going [ɡoʊɪŋ] ←

V

$(\downarrow \text{PRED}) = \text{`GO}\langle(\downarrow\text{SUBJ})\rangle\text{'}$

$(\downarrow \text{ASP}) = \text{PROG}$

$\Downarrow = \downarrow$

Figures 4 and 5 provide extensional visualizations of the structures and relations specified by these lexical entries. The visualization in Figure 4 illustrates that the host must be the subject of the enclitic verb in the functional structure. Figure 5 shows the relations and structures specified by the lexical entry for the verb *going* in (28). These fragmentary lexical structures are merged and integrated in specific syntactic contexts, such as that in Figure 3.

Wescoat shows that the correct f-structure for Figure 3 follows from general principles of structure-function mapping (Bresnan 2001, 103; Bresnan *et al.* 2015). These are visualized in Figure 6; the linking arrows show how the global f-structure corresponds to the c-structure phrases of which D and I are head and co-head, lexically sharing the contraction *you're* which provides their substantive features. (See Wescoat 2005 for more details.)

The main prosodic, syntactic, and morphophonological properties shared by all tensed auxiliary contractions follow from this analysis. (i) Host + auxiliary contractions cannot be conjoined to each other as in (27a) because they are not c-structure units. (ii) The coordination of two auxiliaries together with their verb phrases, despite the first being contracted with the subject as in (27b,c),

---

elements. Specifically, the equation $\Downarrow = \downarrow$ identifies the functional structure of the host *you* with that of the entire contraction, while the equation $(\downarrow \text{SUBJ}) =_c \Downarrow$ imposes the constraint that the host must be the subject of the auxiliary *'re*. To be more precise, it specifies that the f-structure of the contraction (which is identified with that of the atomic host D) must be the value of the SUBJ function of the atomic auxiliary I f-structure. Note that Wescoat (2005) and Wescoat (2009: 612) adopt different but functionally equivalent formulations; the present analysis follows the latter.
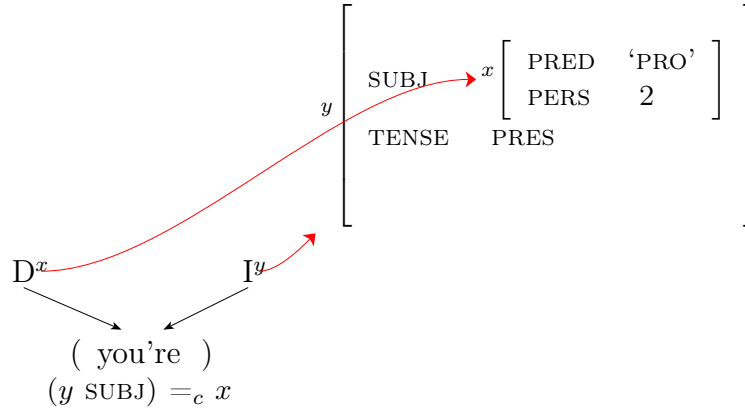
Figure 4: Information specified by the shared lexical entry in (28): the curved arrows represent mappings from c-structure terminals to f-structures and the straight arrows are mappings from the c-structure terminals to their shared lexical exponents.
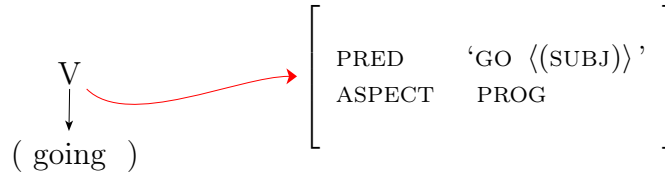

Figure 5: Information specified by the verb lexical entry in (28): again, the curved arrow indicates the mapping from the c-structure terminal to its f-structure and the straight arrow maps from the c-structure terminal to its lexical exponent.

is simply I′ coordination, as Wescoat points out. (iii) The rightward prosodic dependency of the asyllabic auxiliaries matches those of the weak syllabic forms because they are both stressless auxiliary forms occupying syntactically identical positions on the left edge of their complement phrases. (iv) The phonological word status of the host + auxiliary follows from the lexical sharing analysis of tensed auxiliary contractions, given the widely shared assumption of prosodic phonologists that ALL LEXICAL WORDS ARE PHONOLOGICAL WORDS (see, for example, Selkirk 1996).[30]

---

[30]A reviewer points out that contrary to this assumption, Levelt *et al.*'s (1999) lexical access model is designed to allow the phonological word to cross lexical word boundaries. However, their evidence comes from resyllabification between verbs and their unstressed pro-
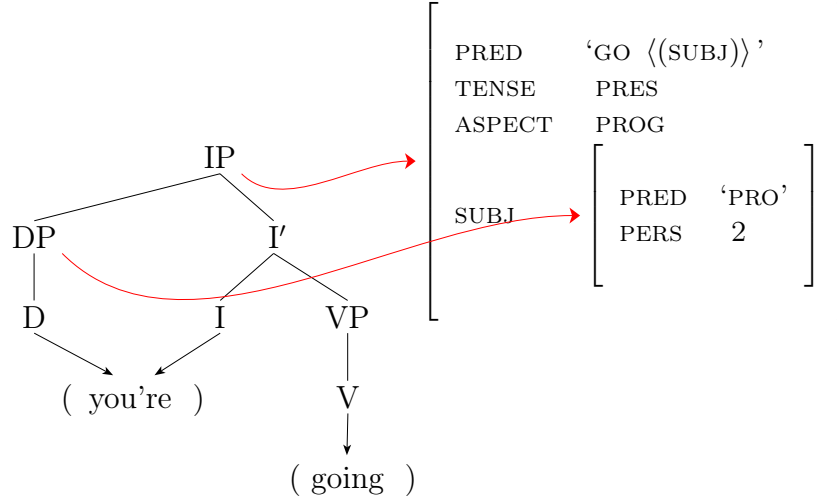
Figure 6: C-structure to f-structure links for the structure in Figure 3 given the lexical entries in (28)

As for the syntactic properties that distinguish restrictive from nonrestrictive auxiliaries (Section 4), those are captured in the formal language allowing reference to the common f-structure of the lexically shared host+auxiliary. Examples occur in the lexical entries making use of the metavariable $\Downarrow$ (n. 29). In (28), for example, the f-structure of the contraction (which is identified with that of the atomic host pronoun D) must be the value of the SUBJ function of the atomic auxiliary I f-structure. This constraint immediately accounts for syntactic restrictions illustrated in (23a–c), where the host cannot be identified with the subject of the verb because it is only part of the subject.

Wescoat broadens the analysis from pronoun subjects to include interrogatives bearing grammaticalized discourse functions (DF) in LFG, and also assumes that the auxiliary may be in its inverted position before the subject (denoted C) as the extended co-head of its clause (Bresnan 2001, 103; Bresnan *et al.* 2015):

noun objects: *escort us* syllabified as es.kor.tus, and *understand it* as un.der.stan.dit (Levelt *et al.* 1999:20, 31). However, there is much evidence that these unstressed pronominal objects in English are not independent lexical words, but enclitics (see, for example, Abercrombie 1961; Selkirk 1972, 1996; Zwicky 1977) so they would not be true examples of resyllabification across lexical word boundaries. Selkirk (1996) analyzes them as "affixal prosodic clitics" Note that while all lexical words are phonological words, some phonological words might be produced from syntactic enclisis (Section 7).

(29) *how've* [haʊv] ←   ADV   C

$(\downarrow \text{PRED}) =$ 'HOW' $(\downarrow \text{TENSE}) = \text{PRES}$

$\Downarrow = \downarrow$ $(\downarrow \text{ASPECT}) = \text{PERF}$

$(\downarrow \text{FOCUS}) =_c \Downarrow$

This extension allows restrictive contractions with interrogative pronouns in a parallel way. The lexical entries allow feature selection of the host by the auxiliary.

Thus the theory of lexically shared clitics adopted here improves on preceding purely prosodic and purely morphological theories of restrictive auxiliary contraction by analyzing them as lexical units whose components simultaneously retain some syntactic independence in c-structure.

6 Lexical sharing of nonrestrictive contractions

Wescoat (2005:482) proposes extending the theory of lexical sharing from restrictive contractions of tensed auxiliaries to the nonrestrictive tensed auxiliary contractions (and indeed to all simple clitics), but he leaves the analysis undeveloped beyond these comments:

> "There is a lexical process that attaches *'s* [z/s/əz] (*is* or *has*) to a host, yielding a lexical-sharing structure; the host may be anything, the attachment of *'s* [z/s/əz] triggers no morphophonological idiosyncrasies, and no functional restrictions are involved. The lack of morphophonological and functional intricacies in no way undermines a lexical-sharing analysis."

It is not difficult, however, to provide a lexical sharing analysis of *'s* contractions. (30) shows the schematic form of lexical entries of contractions of *'s*.[31] It differs from the entry for *you're* shown in (28) in that here the restriction $(\downarrow \text{SUBJ}) =_c \Downarrow$ is absent and the host and its category are unspecified. This schema can be viewed as Wescoat's (2005) "lexical process" for attaching nonrestrictive *'s* to hosts.

---

[31]The generalization to contractions of inverted *'s* would allow C as an extended head as well as I; see discussion of (29).
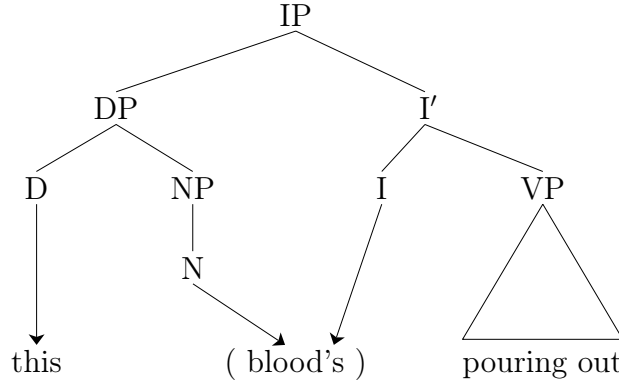
Figure 7: An example c-structure with *'s* contraction under lexical sharing

(30) Schematic form of lexical entries of contractions of *'s*:

$$x\text{'}s\ [\ \ldots\ \text{z/s/ɨz}\ ]\quad \leftarrow\quad \begin{array}{cc} X & I \\[2pt] \ldots & (\downarrow\ \text{TENSE}) = \text{PRES} \\[2pt] \Downarrow = \downarrow & (\downarrow\ \text{SUBJ NUM}) = \text{SG} \\[2pt] & (\downarrow\ \text{SUBJ PERS}) = 3 \end{array}$$

An example of *'s* contraction under lexical sharing is given in Figure 7, and the lexical entry of the contraction *blood's* is given in (31). Note that the lexical entry has the schematic structure in (30), which requires adjacency in c-structure between the host and auxiliary categories. As with other instances of lexical sharing, the host and contracted auxiliary that satisfy the lexical schema form a phonological word.

Figure 8 shows how the structure in Figure 7 corresponds to the global f-structure that results from the same principles of structure-function mapping as before. Under this theory D and NP are co-heads, just as I and VP are co-heads. Because the f-structures of co-heads unify, the features of the NP dominating the host N are unified with the features of the proximate demonstrative D *this*.

(31) Lexical entry for the contraction *blood's* in Figure 7, derived from the schema (30).

$$\textit{blood's}\ [\text{blʌdz}]\quad \leftarrow\quad \begin{array}{cc} N & I \\[2pt] (\downarrow\ \text{PRED}) = \text{`BLOOD'} & (\downarrow\ \text{TENSE}) = \text{PRES} \\[2pt] \Downarrow = \downarrow & (\downarrow\ \text{SUBJ NUM}) = \text{SG} \\[2pt] & (\downarrow\ \text{SUBJ PERS}) = 3 \end{array}$$

PRED   'POUR ⟨(SUBJ)⟩'
TENSE   PRES
ASPECT   PROG

IP

DP   I′

D   NP   I   VP

N

this   (blood 's)   pouring out
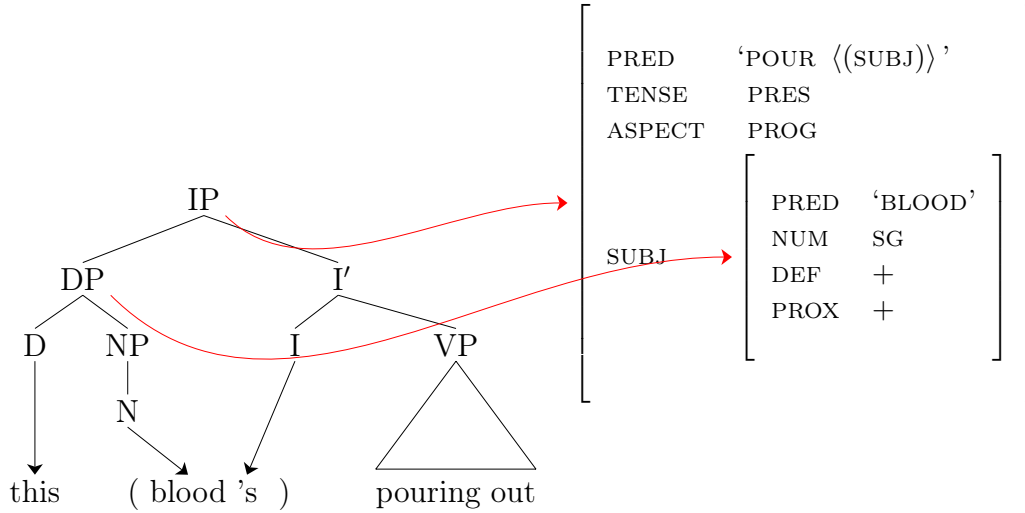
SUBJ

PRED   'BLOOD'
NUM   SG
DEF   +
PROX   +

Figure 8: C-structure to f-structure links for the structure in Figure 7 given the lexical entry in (31) and general LFG principles of structure-function mapping (Wescoat 2005)

A striking property of *'s* contraction, known at least since Baker (1971) and Bresnan (1971), is that *'s* contracts from a sentential complement across a *wh*-extracted subject to a superordinate verb. Examples (32)a–c are authentic examples from the web, selected with negation of the host verb and an affirmative complement in order to eliminate parenthetical readings:

(32) a. *I'll tell you what I don't **think's** going on.* [θɪŋks]

b. *What I don't **think's** beautiful is a boy in my daughter's bedroom.* [θɪŋks]

c. *You can't oppose what you don't **know's** happening.* [noʊz]

As (33) and Figure 9 show, the lexical sharing analysis of these cases is straightforward.

(33)   Lexical entry for the contraction *think's*:

think's [θɪŋks] ←   V   I

(↓ PRED) = 'THINK⟨(SUBJ) (COMP)⟩'   (↓ TENSE) = PRES
⇓ = ↓   (↓ SUBJ NUM) = SG
                                    (↓ SUBJ PERS) = 3

Notice that the lexical entry in (33) is the same in schematic form as that in (31), even though the resulting grammatical relations between the host noun and auxiliary are entirely reversed. To see the reversal, compare Figure 8, where the host heads a subject which is an argument of the main clause co-headed by the tensed auxiliary, to Figure 9, where the host heads the main clause and the tensed auxiliary co-heads a complement clause which is an argument of the host predicate. No special stipulations of functional annotations are required to derive the correct f-structures. Both structures satisfy the adjacency requirements of the schema for nonrestrictive contractions in (30) and follow from the general principles of structure-function mapping invoked by Wescoat (2005).
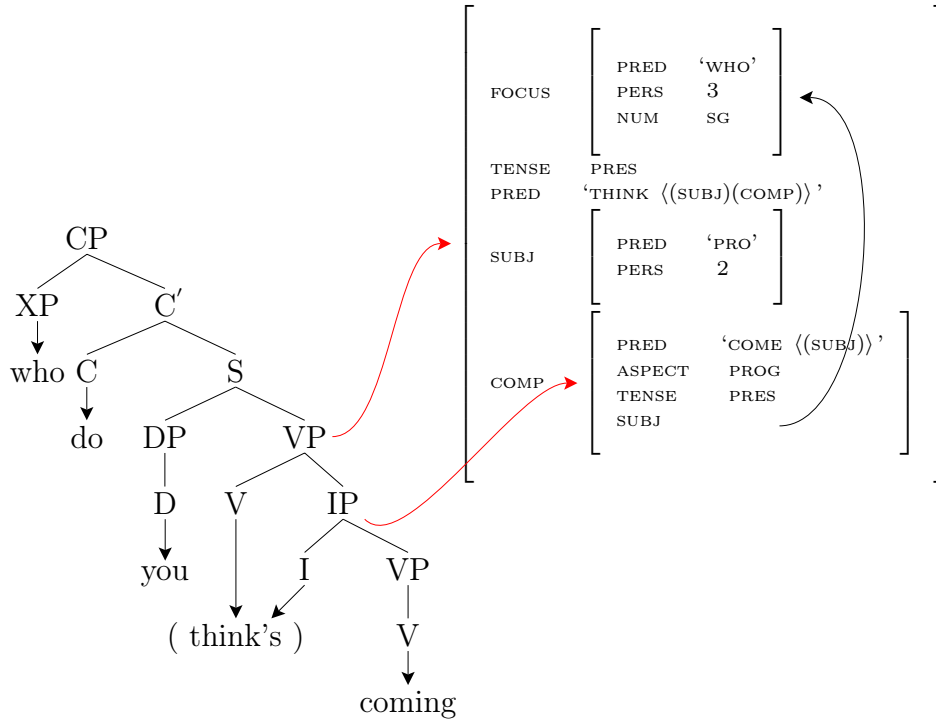

Figure 9: C-structure to f-structure links for a structure using the lexical entry in (33)

Furthermore, since *is* contractions are not c-structure constituents under lexical sharing, there is no danger of unwanted 'movements' in the lexical sharing analysis of nonrestrictive contractions (cf. Anderson 2008, 174): ***Who's** do you think coming?*, cf. *Who do you think is/'s coming?*, ***Who'd** would you*

say accept? vs. *Who would you say would accept?*, and **It's **you're** that sick* vs. *It's you that are sick.*

In sum, the lexical syntactic analysis of tensed auxiliary contractions adopted here not only improves on previous accounts, but extends gracefully beyond them in empirical coverage.

7   A hybrid model

Combining the formal grammar and the usage-based mental lexicon reviewed in previous sections into a hybrid model is the subject of the present section. The aim is not to present a detailed formalization, but to describe at a high level how the architecture of the dynamic exemplar model discussed in Section 3 could combine with the formal grammar of the present study to explain the main empirical findings of both usage-based and formal lines of research on tensed auxiliary contraction.

In the present framework a hybrid model of SYNTACTIC production (excluding higher-level discourse context and semantics) would use f-structures as input representations, lexical entries as labels of memory-trace clouds, and the ordered lexical exponents of c-structures as outputs. These concepts are illustrated in Figures 10 and 11 for the production of *you're* or *you are*.

Figure 10 illustrates an input to speech production at the syntactic level as an abstract plan for a phrase or sentence. The plan is represented by a functional structure for a second person pronoun subject of a clause in the present progressive. Activation of this f-structure would activate the words that are linked to it in the mental lexicon: *you're*, *you*, and *are*. These are the labels most similar to the input in their relational features—specifically, the words whose functional schemata in their lexical entries can be instantiated to match the input f-structure. (Compare the extensional visualizations of the functional schemata of lexical entries in Figures 4 and 5.)

These lexical entries would each label a cloud of memory traces, like the illustration in Figure 2, which uses orthographic words as labels. The word clouds of *you* and *are* would be bound together by their links to the same input f-structure and as a set would serve as a composite label for union of the word clouds for *you* and *are*. Thus the hybrid model incorporates both contractions and their uncontracted multiword equivalents in the mental lexicon (cf. Section 3).

An exemplar would be randomly selected as a target of production from

the union of the clouds of *you're* and the composite label *you are*. If nothing differentiates them in the input grammatical context, the contracted and uncontracted variant exemplars would both be possible selections as targets of production.

$$
\begin{array}{ll}
x &
\begin{bmatrix}
\text{SUBJ} & y\begin{bmatrix}\text{PRED} & \text{`PRO'} \\ \text{PERS} & 2\end{bmatrix} \\
\text{TENSE} & \text{PRES} \\
\text{ASPECT} & \text{PROG}
\end{bmatrix}
\end{array}
$$

D^y I^x

(you're)
$(x \text{ SUBJ}) =_c y$
[jʊɹ] [jɔɹ]
[juːɹ]
[jʊɹ]
[juːɹ]
⋮

D   I

you   are
[juː]   [ɑɹ]
[jə]   [əɹ]
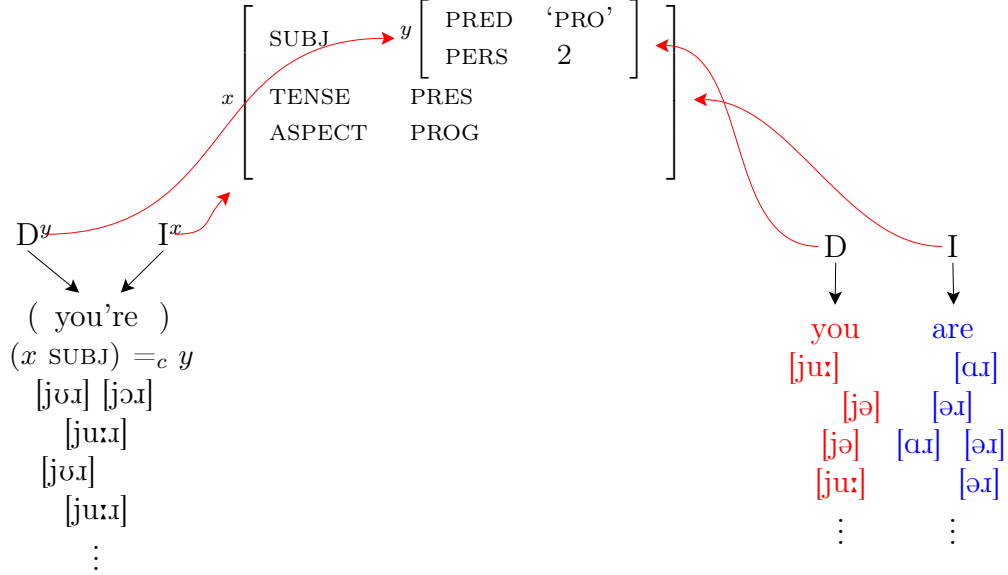[jə]   [ɑɹ] [əɹ]
[juː]   [əɹ]
⋮   ⋮

Figure 10: A visualization of a production input as an abstract phrase or sentence plan (an f-structure) linked to words in the mental lexicon whose functional schemata match it. The words label clouds of memory traces, from which an exemplar is randomly selected as the target of production.

To produce a syntactic output from the randomly selected production target, the syntactic production process would fit the winning exemplar into the phrase patterns of English in accordance with its lexical entry or entries so that it corresponds to the input f-structure. Details of generation and parsing are outside the scope of the present study,[32] but the syntactic output of the example input could be one of the alternative strings of ordered lexical exponents in the c-structures shown in Figure 11.[33]

---

[32] Wedekind & Kaplan (2012) discuss various computational linguistic generation algorithms for LFG.

[33] The curved arrow mappings from IP and DP to the f-structure in Figure 11 arise because in the syntax the f-structure of a node is identified with that of its head or co-head (Bresnan *et al.* 2015). Alternative theories of c-structure could of course be adopted, with
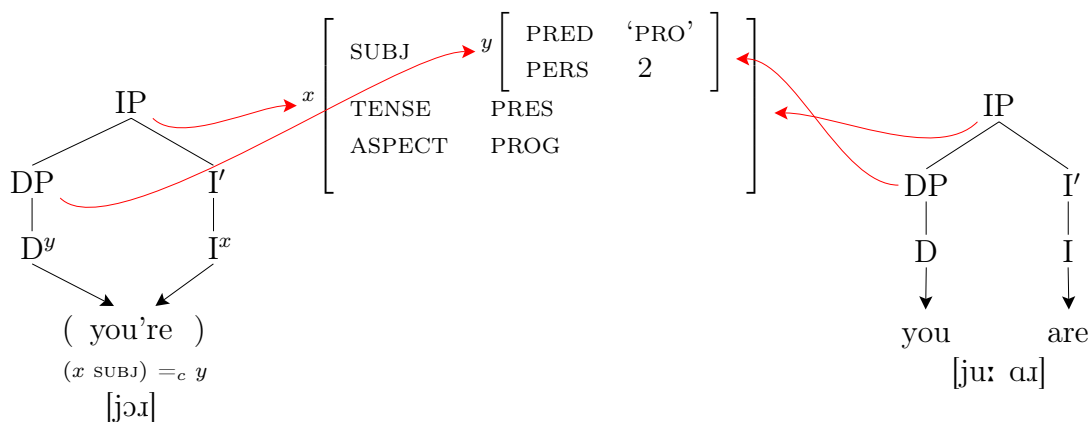
IP $\xrightarrow{x}$

$\left[\begin{array}{ll} \text{SUBJ} & y\left[\begin{array}{ll}\text{PRED} & \text{'PRO'} \\ \text{PERS} & 2\end{array}\right] \\ \text{TENSE} & \text{PRES} \\ \text{ASPECT} & \text{PROG}\end{array}\right]$

DP   I′

D^y I^x

(you're)

$(x\ \textsc{subj}) =_c y$

[jɔɹ]

IP

DP   I′

D   I

you   are

[juː ɑɹ]

Figure 11: A visualization of alternative production outputs

Figures 10 and 11 represent a synchronic model of production, but the diachronic applications of the dynamic exemplar model elsewhere (Section 3) lead to the question of why the syntactic structure on the left hand side of Figure 11 arises as a variant of that on the right.

Observe that the contraction cannot simply be a sequence of phonetically fused words or allomorphs of the adjacent pronoun *you* and the auxiliary verb *are*, as described in Section 3, because the fusion does not occur everywhere that the sequence occurs. Recall (23b–c), for example [*The people beside* **you**]*'re going*, pronounced [juː..ɹ] but not [*jʊɹ/*jɔɹ]. Thus what is lexically stored is not merely a sequence of words and allomorphs, but fragments of syntactic structures they occur in with their local relations and dependencies, as visualized in the left side of Figure 11. These syntactic fragments can enter into conjuncts parallel to uncontracted phrases, as in **You**[*'re gonna do what I say*] *or* [*will suffer the consequences*] (cf. (27b,c)). And they share the rightward metrical dependence of unstressed *are* in uncontracted phrases (Section 4).

Consequently, at the syntactic level the lexical storage of high-probability restrictive auxiliary sequences like *you* and *'re* as units must include the storage of the fragments of syntactic structure they occur in. This is what lexical sharing does: it specifies the contracted sequence *you're* as a sequence of word categories that share a common functional structure in which *you* is the required subject of *are*.

varying degrees of flatness or hierarchical structure and finer or coarser-grained part-of-speech categories (n. 28).

Observe that the syntactic restriction of the asyllabic auxiliary *'re* to subject pronoun hosts singles out the syntactic context that has the highest share of token frequencies of cooccurrence with the auxiliary.[34] For this reason, lexical sharing as a formal construct can be viewed as a grammaticalization of high-probability syntactic distributions in usage just as the allomorphs of *you* and unstressed *are* can be seen as a grammaticalization of high-probability pronunciations.

How should the model handle *is* contractions? With restrictive auxiliaries like *are, 're* in Figure 10, the high probability of cooccurrence with their pronoun hosts leads to repeated phonetic reductions of the host which become lexicalized over the long term, providing independent support for the lexical storage of the host + auxiliary combinations as units. But with the nonrestrictive auxiliary *'s*, evidence of such long-term phonetic reductions of hosts is lacking. At the same time, authentic examples like (24a–d) and (32a–c) suggest that this auxiliary lacks all syntactic constraints on its host except adjacency (but see Section 9). The schematic shared lexical entry for *'s* (30) expresses both of these properties: it neither selects a specific lexical host nor imposes the requirement that the host be its subject or have any relation other than being an adjacent word category to the left. Imported into the model of the mental lexicon, this entry would essentially provide a lexical label for the clitic *'s* without a specific host, simply as an allomorph of *is*.

In the mental lexicon, the clitic *'s* would label a cloud of memory traces just as the uncontracted *is* does (cf. *are* in Figure 10). Then its activation, selection, and output production would proceed like that of *is*, except that its lexical entry would specify an adjacent host of any category to its left. The output production process would cliticize *'s* onto its host in accordance with its lexical entry (30), forming a phonological word (cf. Inkelas 1991, Inkelas & Zec 1993), and then fit it into the c-structure patterns of English that correspond to the input.

By itself, this analysis of contracted *'s* would yield a free and productive choice of *'s*, like *is*, for any adjacent host. Productions of *is* contraction could take place with novel hosts. And if that were the whole story, the probability

---

[34]The reasoning for this claim is that subject pronouns far outnumber non-pronouns adjacent to the auxiliary (Section 8), and with respect to the ungrammatical cases (23) each additional level of structural embedding in the syntax of the host phrase introduces other possible heads which impose alternative selectional restrictions to those the auxiliary itself, serving to increase the type frequency and decrease the token frequency of words in that host position.

of *is* contraction would be independent of the word serving as host. Instead of being conditioned on the joint occurrence of host with *is* forms, it would be roughly constant across non-pronoun hosts, dependent only on the proportions of the clitic *'s* and the syllabic forms of *is*.

However, if speakers produce the host word adjacent to the clitic *'s* sufficiently often, the sequence could become a lexically stored unit, parallel to *you're* in Figure 10. The assumption needed for unit formation to occur is the perception-production loop: what is produced is perceived and stored and that will include generated productions. Given memory decay, infrequent and temporally remote stored combinations would become inaccessible as units and require generation by cliticization. In contrast, frequent and recent composite exemplars like, say, *Mum's*,[35] could become increasingly accessible established units. In this way, *is* contractions could in principle have dual sources either as stored units with specific hosts or as freely generated cliticizations,[36] and would show increasing contraction with sufficiently increasing frequencies of cooccurrence of host and auxiliary.

In sum, the hybrid model incorporates the usage-based explanation for the fact that the frequency of cooccurrence of host + auxiliary correlates with their likelihood of contraction (Section 3). But because the labels of its exemplar clouds are lexical and lexically shared representations of formal grammar that have well-defined mappings to syntactic input and output structures, it also entails the grammatical properties that restrictive auxiliaries share with equivalent uncontracted phrases (Sections 4–5). Hence, the hybrid model has broader explanatory scope than either of its usage-based or formal-grammar based components alone.

8   A corpus study

The analysis of *is* contraction in the preceding section suggests alternative predictions about the probability of contraction with non-pronoun hosts. Under the traditional generative analysis, contracted *'s* is simply a clitic allomorph of *is* not stored with its host, but generated by a cliticization process in the production of outputs. Under this analysis the probability of *is* contraction would be independent of the word serving as host. Under the alternative analysis

---

[35]As the next section shows, *Mum* is one of the most frequent nouns that occurs before *is* or *'s* in the Canterbury Corpus.

[36]Cf. Lowe's (2016) lexical sharing analysis of genitive *'s*, n. 54.

provided by the hybrid model, sufficiently high-probability host + auxiliary sequences would achieve persistent storage as units, leaving only very infrequent combinations for generative cliticization. Under this alternative analysis, the probability of *is* contraction would be higher with highly frequent host + auxiliary sequences.

To investigate these theoretical predictions, *is* contraction data were collected from two spoken corpora, the Buckeye Corpus (Pitt *et al.* 2005) and the Canterbury Corpus (Gordon *et al.* 2004). (Nonrestrictive *'d* contractions are set aside because they are much sparser; recall the discussion of (25).) The Canterbury Corpus is over three times as large as the Buckeye Corpus and is annotated for the social variable of class that could play a role in contraction (cf. Labov 1969, McElhinny 1993).[37] Hence, it became the main focus of the investigation, with the Buckeye Corpus used to replicate the findings from the Canterbury Corpus.

All instances of *is* and orthographic *'s* were collected from the Canterbury Corpus transcriptions in 2015 at the New Zealand Institute for Language, Brain, and Behavior; research assistant Vicky Watson manually checked a sample against the audio files for transcription accuracy and also marked data exclusions. Exclusions included hosts with final sibilants (which do not occur with the asyllabic auxiliary *'s*), instances of *'s* representing *has*, and the variety of other grammatical contexts found by MacKenzie (2012:65–90) to be outside the envelope of variation. The hosts in this dataset were labeled as pronouns or non-pronouns, and informativeness was calculated as in (1) from ngram statistics provided by Jen Hay and Robert Fromont for the entire Canterbury Corpus. This yielded 11,719 total observations from 412 speakers (mean instances per speaker = 28, standard deviation = 23) and 758 unique non-pronoun hosts.

For the Buckeye Corpus replications a dataset of variable *is* contractions was extracted and annotated following a similar method to that of Bresnan & Spencer (2012) and J. Spencer (2014).[38] The author labeled the hosts of this dataset as pronouns or non-pronouns, and calculated informativeness as in (1) from ngram statistics compiled from the entire Buckeye Corpus. After exclu-

---

[37]The Canterbury Corpus has been used primarily for sociophonetic studies and previous studies of auxiliary contraction in NZ English are lacking.

[38]Bresnan & Spencer (2012) and J. Spencer (2014) already show an effect on contraction of (respectively) the log and negative log conditional probability of non-pronoun hosts given *is/'s* in data collected from the Buckeye Corpus. The present dataset was constructed independently of the datasets described in those studies and encompasses a greater range of host phrase lengths.

sions, there were 4019 total observations from all 40 speakers (mean instances per speaker = 100, standard deviation = 46) and 306 unique non-pronoun hosts.

First, non-pronouns have higher informativeness than pronouns before the tensed auxiliary *is/'s*,[39] so on the theory of lexical sharing in the mental lexicon, their likelihood of *is* contractions should be lower. The data bear out this expectation:

- Out of 11,719 total observations of variable full and contracted *is*, 88% follow adjacent subject pronouns and 12% follow adjacent non-pronouns.

- Contraction appears with 96% of the former and 43% of the latter observations.

Comparable data from the smaller Buckeye Corpus of spoken mid-American English show a similar pattern:

- Out of 4019 total observations of variable full and contracted *is*, 85% follow adjacent subject pronouns and 15% follow adjacent non-pronouns.

- Contraction appears with 92% of the former and 37% of the latter observations.

Secondly, among non-pronoun hosts before *is/'s*, those that have lower informativeness should tend to have higher chances of contraction. This expectation is also borne out by data from the Canterbury Corpus. The non-pronoun hosts having lowest informativeness in the Canterbury Corpus *is*-contraction dataset are *one, mum, dad*, and *thing*. These have a far higher proportion of contractions than the average for non-pronouns: 0.837. Some authentic examples appear in (34):

(34)　*and my poor **Mum's** here going oh I wish I was there*

　　　*and I said come quick come quick . **Dad's** at home and he's a hell of a mess*

　　　***one's** a um . a raving . feminist an **one's** a chauvinist*

---

[39]This observation immediately follows from the fact that pronoun-*is/'s* bigrams are far more frequent than non-pronoun-*is/'s* bigrams and the definition of informativeness in (1), given n. 7.

*I've got [a] friend that has three cats and **one's** a really spiteful cat .*

*liturgy that they all join in on . and the whole **thing's** sung*

*I wonder if that that kind of **thing's** like hereditary*

In the Buckeye Corpus *is*-contraction dataset, the proportion contracted of the least informative non-pronoun hosts (*everybody, one, think, everything, daughter*, and *mom*) differs less extremely from the other hosts, but shows a tendency in the predicted direction: 0.453 vs. 0.354.

These simple descriptive statistics support this crucial consequence of the hybrid theory: that *is* contraction with non-pronoun hosts should show evidence of the probabilistic structure of the mental lexicon. But while these data points are suggestive, what is needed to test the prediction is a statistical model that controls for other possible predictors of contraction. After all, there are many hosts in the dataset, and the literature on contraction has identified many contributors to *is* contraction other than informativeness (see below). To this end, the author planned a multiple logistic regression model and annotated the corpus data for the variables described below, using the statistical computing platform R (R Core Team 2019) as well as direct inspection and manual annotation of extensive data samples.

Informativeness

The main variable of interest, the informativeness of the non-pronoun host before *is/'s*, is calculated as in Section 2. Here, the estimates of bigram and unigram probabilities come from the frequencies of host + *is/'s* and *is/'s* in the entire Canterbury Corpus of 1,087,113 words.

Host phrase WC

Host phrase word count (WC) is one of the best predictors of contraction (Frank & Jaeger 2008; MacKenzie 2012, 2013; Bresnan & Spencer 2012; J. Spencer 2014). WC can be viewed as a convenient proxy for phrasal weight or complexity, which may make the host phrase more likely to be phrased separately, set off by a phonological or intonational phrase boundary.[40] It could also be viewed as a proxy for phrasal informativeness, in that longer phrases are likely

---

[40]See Szmrecsányi (2004) on operationalizing syntactic complexity.

to be more informative in a qualitative sense.[41] Table 5 provides authentic examples; Table 6 shows the relation of word counts (excluding pronoun hosts and counting space-separated character strings by an R script) to contractions in the data.

host phrases (bolded)	word count
*but now **work**'s just so busy ...*:	**WC**=1
***the work**'s so much harder*:	**WC**=2
***all this blood**'s pouring out the side of my head*:	**WC**=3
***some of the work** is a bit tedious*:	**WC**=4

Table 5: Host phrase word count (WC)

host phrase WC:	1	2	3	4+
total instances:	543	500	182	143
proportion contracted:	0.74	0.55	0.43	0.18

Table 6: Proportion contracted by host phrase word count

Year of birth

With non-pronoun hosts, younger speakers of New Zealand English (those born from 1961 to 1987) use contraction more than older (those born from 1926 up to 1961), as Table 7 shows.

year of birth:	[1926,1961)	[1961,1987]
proportion contracted:	0.50	0.61

Table 7: Proportion contracted by speaker year of birth

Speaker year of birth is numerical data available in the corpus, but it is severely bimodal around the year 1961, causing model fit problems. The year of birth data is therefore dichotomized at 1961.

---

[41]However, quantitative measures of phrasal informativeness run up against the problem of sparseness of data. Even restricting host phrase length to two words, for example, one finds that 90% of the 500 two-word phrases occur just once in the dataset.

Class

Nonprofessional NZE speakers use contraction more than professionals (Table 8).

class:	N	P
proportion contracted:	0.65	0.49

Table 8: Proportion contracted by speaker class

Previous instance

If the previous instance is *is* or *'s*, the likelihood of *is* contraction is respectively lowered or raised (Table 9). See Szmrecsányi (2005) on "structural persistence."

previous instance:	*'s*	*is*	none
proportion contracted:	0.599	0.342	0.510

Table 9: Proportion contracted by previous occurrence of *is/'s*

Successive instances of *is/'s* are from the same speaker, are collapsed across the copula/auxiliary types (see below), and include all previous contractions, including those with pronoun hosts.

_is_ type

Those instances of *is/'s* in construction with a participial form of the verb are defined as 'auxiliaries', while those in construction with nominals, prepositions, and adjectives are defined as 'copulas'. The *is* auxiliary verb contracts more than the *is* copula (cf. Labov 1969; Rickford *et al.* 1991; McElhinny 1993; MacKenzie 2012; J. Spencer 2014), as Table 10 shows.[42]

is type:	aux	cop
proportion contracted:	0.635	0.548

Table 10: Proportion contracted by auxiliary type

---

[42]For a more refined analysis of construction types see Barth (2011) and Barth & Kapatsinski (2017) and also compare MacKenzie's (2012) discussion of following constituent category.

Other predictors

Other potential predictors were considered for inclusion: speaker's gender, whether the final segment of the host is a consonant or vowel, the stress level of the final segment, the length of the host in segments, and the number of syllables of the host. All of these added nothing to the model: they had coefficients less than the standard error and were dropped. Interactions were not included because of the complexity of the model in relation to the data.

In addition, various metrical or prosodic properties of the hostphrase were tested as alternatives to WC for another project: (1) total metrical feet (Sternberg *et al.* 1978, Sternberg *et al.* 1988); (2) edge boundary strength, manually annotated as the number of lexical word brackets summed with the number of major syntactic phrase (NP, VP, CP) brackets that separate the host from the verb, theoretically corresponding to phonological phrases in Match Theory (Selkirk 2011); (3) cumulative stress from manual annotation of perceived stress values, with and without transformation to a grid format (Liberman & Prince 1977); and (4) cumulative stress based on manually corrected automatic annotation of theoretical stress values, transformed to grid formats. (1) and (4) were automatically annotated using software developed by Anttila *et al.* (To appear). WC substantially improves the model fit compared to alternatives (1) and (4), while (2) and (3) are both competitive with WC. WC is retained here as a convenient proxy pending further research.

The fitted model

Because speaker identity is a source of unknown dependencies in the data, a multiple logistic regression "working independence" model (Harrell Jr 2001) was constructed from these variables, with the numerical variables standardized.[43] After the model was fitted to the data, it was corrected for intra-speaker cor-

---

[43]Here the working independence model starts from the assumption that speakers' utterances are independent of speaker identity, and then corrects this assumption by estimating the extent of these dependencies using bootstrap resampling with replacement of entire clusters (each speaker defines a 'cluster' of utterances). Bresnan *et al.* (2007) describe cluster resampling in this way: "In other words, we can create multiple copies of the data by resampling from the speakers. The same speakers' data can randomly occur many times in each copy. We repeatedly re-fit the model to these copies of the data and used the average regression coefficients of the re-fits to correct the original estimates for intra-speaker correlations. If the differences among speakers are large, they will outweigh the common responses and the findings of [the working independence model] will no longer be significant."

relations by bootstrap cluster sampling with replacement using the bootcov() function of Harrell Jr (2018). The resulting parameter values are shown in the final fitted model in Table 11.

$$\text{Prob}\{\text{Contracted} = 1\} = \frac{1}{1 + e^{-X\beta}}, \quad \text{where}$$

$$X\hat{\beta} =$$
$$0.8804$$
$$-0.4741 \times -log_2 P(\text{host}|\text{verb})$$
$$-0.9868 \times [\text{previous instance} = is]$$
$$-0.2177 \times [\text{previous instance} = \text{none}]$$
$$-1.0068 \times \text{host phrase WC}$$
$$-0.7060 \times [\text{class } = \text{P}]$$
$$-0.5370 \times [is \text{ type } = \text{cop}]$$
$$+0.4515 \times [\text{year of birth} = [1961, 1987]]$$

and $[c] = 1$ if subject is in group c, 0 otherwise

Table 11: Model of Canterbury Corpus variable *is* contraction data with non-pronoun hosts

The model in Table 11 predicts the probability of contraction of any example, given its predictor values. The top line formula converts log odds (used by the regression model) to probabilities. Below, the initial value 0.8804 is the INTERCEPT, representing the overall likelihood of contraction (measured in log odds) when all of the predictor values are zero. The subsequent numerical values are COEFFICIENTS of the model formula, which weight the various predictors and show whether they increase or decrease the overall log odds of a contraction when they do not have zero value; positive coefficients add to the likelihood of the contraction given by the intercept, while negative coefficients reduce the likelihood. The predictors in square brackets are binary-valued indicators of categorical properties—professional/nonprofessional *class*; auxiliary/copula *'is' type*; speaker *year of birth* in the earlier or later interval of years. One of the categorical property values is taken to be zero and included in the intercept to calculate the overall likelihood of contraction; when the alternative property

43

value is observed, the overall likelihood is accordingly adjusted by multiplying the coefficient by 1 and adding the result to the total.[44] The non-categorical predictors $-log_2 P(\text{host|verb})$ (informativeness of the host given the verb) and host phrase WC (host phrase word count) have scalar values which are also multiplied by their coefficients. This and similar model formulas are used to validate the model by assessing its predictions on unseen data.

The model quality is reasonably high.[45] Partial effects of the model are plotted in Figure 12. The predictors are all reliable within 95% confidence bands, except for the case when the value of *previous instance* is "none"; there were too few data points for that estimate to be reliable. Because the scalar predictors are standardized, they are plotted on the same scale and the much larger effect of host phrase WC is clearly visible from the greater range it covers on the *y*-axis. The informativeness of the host nevertheless has a clear effect as well: greater informativeness depresses the log odds of contraction.

The author replicated this finding on the non-pronoun host data from the Buckeye corpus of spoken mid-West American English. The predictors are the same except for age and class, which were unavailable or unrelated to contraction in this dataset. Modeling and validation by the same methods as before showed a reliable effect of informativeness of the host on contraction.

Barth & Kapatsinski (2017:40–41) conducted a multi-model analysis of *is/'s* contractions with non-pronoun hosts in a smaller dataset of spoken language from the Corpus of Contemporary American English (Davies 2008–). They report that by far the most explanatory predictor among those they used is the bigram probability of host (their "Preceding JP" and Krug's 1998 "string frequency"), which is proportional to the informativeness of the host (n. 7).

In sum, this prediction of a hybrid theory has been borne out by empirical studies of two spoken English corpora in the present study and is buttressed by a further empirical study of a third corpus (Barth & Kapatsinski 2017): usage probabilities affect not only the contractions of restrictive auxiliaries with their pronoun hosts and morphophonological fusions, they also affect in the same way the contractions of the most unrestrictive auxiliary *is* with noun hosts.

---

[44]The three-valued predictor for *previous instance* is decomposed into two binary two-valued predictors: full *is* vs. *'s*, and no previous instance vs. *'s*.

[45]Validation of the model found that more than 95% of averaged observed minus expected values in 35 bins are within 2 standard errors (see Gelman & Su's 2018 binnedplot() function); all predictors have low multicollinearity (condition number $c < 5$, vif < 1.1); average Concordance is $C > 0.758$ under 10-fold cross-validation with bias correction for speaker clusters in each fold—an "optimism" of < 0.01.
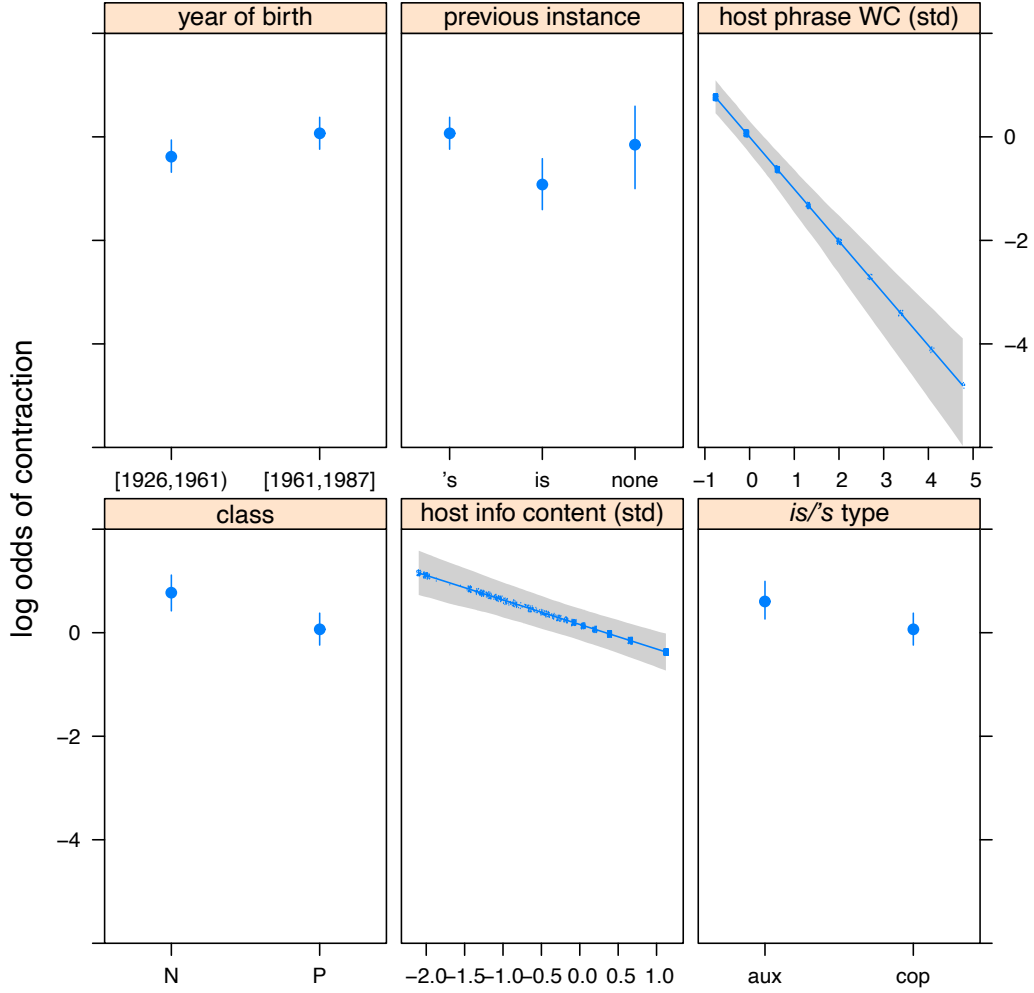
Figure 12: Partial effects of the model in Table 11. Each panel shows the effect of one predictor when all of the others are held constant. 95% confidence bands are from the bootstrapped cluster resampling of speakers. The vertical ticks on the plot lines of the numerical covariates (host phrase WC and host informativeness) show the data densities along the predictor scales.

Given that the variable of interest is informativeness, a measure of usage probability, and none of the other predictors in the model specifically relates to

the formal grammar of *is* contractions discussed in Section 4, what empirical contribution does the formal grammar make to the corpus study? The answer is that the evidence for formal grammar was carved out of the collected data in advance of modeling, in order to ensure that the remaining dataset was within the "envelope of variation" defined by MacKenzie (2012). The excluded data were cases where contractions with asyllabic auxiliaries are blocked by known grammatical factors.[46] These grammatical predictors of blocked contraction create near-perfect separations of the output (contracted/uncontracted) at the population level, meaning the level of the whole language from which the corpora are sampled; therefore the mathematics of logistic regression cannot be applied to estimate their probabilities.

9   Between restrictive and nonrestrictive

As Section 7 points out, the restrictive auxiliaries' syntactic restrictions to subject pronoun and interrogative pro-form hosts single out the syntactic positions that have the highest token frequencies of cooccurrence with the auxiliaries (cf. (23) and n. 34). The formal descriptions of these syntactic restrictions in the lexical entries of auxiliaries in Sections 5 and 6 can then be regarded as describing grammaticalizations of distributional usage patterns. A closer examination reveals that as one might expect from the grammaticalization of usage patterns, the line between restrictive and nonrestrictive auxiliaries is not a binary categorical classification as implied in Section 4.

Although the restrictive auxiliary *'ve* overwhelmingly occurs with subject pronouns and interrogative pro-forms, with low probability it does contract with some host nouns, such as example (35) (Barron 1998:247, n. 13)—

(35)   *The **BBC'**ve reported* ... [biːbiːˈsiːv]

—and the following example from the Buckeye corpus:

(36)   *... all their life **people'**ve been saying...* [ˈpiːpl̩v]

---

[46]—such as clause-final and phrase-final occurrences of the auxiliary, *is* in specification constructions like (14), and pauses or adverbs intervening between host and auxiliary, as well as negated auxiliaries (*isn't, is not*) and hosts ending in final sibilants. Various errors of transcription were also excluded.

Further blurring the boundary between restrictive and nonrestrictive auxiliaries, there are subtle differences in selectivity for hosts among the nonrestrictive auxiliaries. For example, *'s* for inverted *is, has* contracts with all *wh*-proforms,[47] but (at least in the author's speech) *'d* for *had, would* doesn't. *'d* contracts with *who* (*Who'd like to come with me?* and *someone who'd failed*) but not *how*, unless it represents inverted *did*:

(37)   *How's it going?* [haʊ.əz, haʊz]   *'s < is*
        How's it gone so far? [haʊ.əz, haʊz]   *'s < has*
        How'd it happen? [haʊd]   *'d < did*
        *How'd it happened? [*haʊd]   *'d < had*
        *How'd it have happened? [*haʊd]   *'d < would*

As mentioned in Section 2, even nonrestrictive *'s* has a number-neutral use where it selects for a small set of pro-form hosts allowing both singular and plural complement nouns, unlike the full form *is* (Dixon 1977, Nathan 1981, Sparks 1984, Kaisse 1985):

(38) a.   *Where's my pants?*
        *In what location's my pants?
        (cf. **Where is my pants?*)

    b.   *How's your feet?*
        *In what condition's your feet?
        (cf. **How is your feet?*)

    c.   *There's the cattle.*
        *The cattle's there.
        (cf. **There is the cattle.*)

These data indicate that there are intermediate usage patterns between the restrictive and nonrestrictive types presented in Section 4.

Formalizing the lexical entries for these intermediate cases provides a more systematic picture of their grammar. For example, the number-neutral use of *'s* illustrated in (38a–c) can have lexical entries similar to (39):

---

[47]Kaisse (1983) makes the interesting observation that inverted *is* contraction is more restricted than *is* contraction with the subject: *Which dog's been jumping on the sofa?* (subject) vs. **What dog's that?* (inverted with subject). Inverted *is* contracts with an interrogative pro-form itself (*What's that?*) but much more rarely with a host embedded in an interrogative phrase. Judgments are uncertain, but could indicate a usage probability effect for inverted *is*, like that in n. 34.

(39) *how's* [haʊz] ← ADV C
$\quad\quad$ (↓ PRED) = 'HOW'$\quad$ (↓ TENSE) = PRES
$\quad\quad\quad\quad$ ⇓ = ↓ $\quad\quad\quad$ (↓ SUBJ PERS) = 3
$\quad\quad\quad\quad\quad\quad\quad\quad\quad$ (↓ FOCUS) =$_c$ ⇓

Unlike the syllabic auxiliary forms, the asyllabic auxiliary specifies person but not the number of the subject and it selects for specific pro-forms as hosts *how*, yielding *How's your feet?* vs. **In what condition's your feet?* and **How is your feet?*

As also noted in Section 4 the asyllabic auxiliary *'d* is restrictive in some varieties and nonrestrictive in others. A restrictive entry for the conditional mood sense of *'d* requiring a pronoun subject is shown in (40):[48]

(40) *I'd* [aɪd] ← D I
$\quad\quad$ (↓ PRED) = 'PRO'$\quad$ (↓ MOOD) = COND
$\quad\quad$ (↓ SUBJ PERS) = 1$\quad$ (↓ SUBJ) =$_c$ ⇓
$\quad\quad\quad\quad$ ⇓ = ↓

In this variety *Bligh'd have seen it* pronounced [*blaɪd] is ungrammatical. In other varieties the pronoun specification on the host is dropped and the subject [blaɪd] is fine:

(41) *x* [...d] ← X I
$\quad\quad\quad$ ⇓ = ↓ $\quad$ (↓ MOOD) = COND
$\quad\quad\quad\quad\quad$ (↓ SUBJ) =$_c$ ⇓

Because the host f-structure must be identified with that of the subject of the auxiliary, this shared entry rules out the contraction of conditional or past-perfect *'d* with adverbs, as in examples like *So'd Ann* *[soʊd] for *So would Ann*, as well as accounting for **How'd it have happened?* in (37).

For the present author *'d* is even less restrictive, allowing contractions not only with a subject as in (41), but with an adjacent dependent of the subject: witness *family'd* in (26). The greater degree of contraction is permitted by the lexical entry in (42):[49]

---

[48]Recall from (37) that past perfect and conditional uses of *'d* differ in host selectivity from the past tense use.

[49]The notation GF* specifies a possibly empty chain of nested grammatical functions, allowing nonlocal dependencies between the auxiliary's subject and its host. For this and other details of the formalism see Börjars *et al.* (2019) or Dalrymple *et al.* (2019).

(42) x [...d]   ←   X   I
                        $\Downarrow = \downarrow$ ($\downarrow$ MOOD) $=$ COND
                                      ($\downarrow$ SUBJ GF*) $=_c \Downarrow$

The same shared entry rules out *Who would you say 'd accept?* because the
host *say* of contracted *'d* is not a dependent of the subject of *would*.

It is plausible that usage probabilities underlie these specific differences in
syntactic distribution, shaping the synchronic grammar of auxiliaries as they
have their diachronic development (Bybee 2010). But with the relative paucity
of *'d* contractions in corpora, and the infrequency of long host phrases in spo-
ken language in general (n. 41), the necessary research would probably require
experimental methods beyond the scope of the present study.[50]

The formal syntactic analyses illustrated above also suggest a path by which
auxiliaries can change from one type to another: it is by a kind of "syntactic
bleaching" in which relational features are gradually lost, initially by becoming
optional, which reflects variable restrictiveness, and eventually by dropping the
feature option altogether. The auxiliaries in the respective entries (40) and (41)
for the British and American varieties differ by the loss of the feature specifying
a pronoun host. A fully unrestricted *'d* parallel to the unrestricted *'s* in (30)
would differ from both (41) and (42) by the loss of the feature constraining the
host to be the subject of the auxiliary. A rich lexical syntactic literature on the
development of agreement markers from pronoun clitics in multiple languages
(see Bresnan *et al.* 2015, ch. 8 and references there) shows that this kind of
feature optionality and loss is a natural progression which is well captured by
the relational specifications of the formal grammar.

10   *I dunno* parallels and implications

The formal theory of lexically shared host +auxiliary contractions extends fur-
ther into the larger domain of multiword expressions, such as Bybee & Scheib-
man's (1999) study of *I don't know* discussed in Section 1. The formal analysis
brings out parallels between this multiword expression and the grammar of
tensed auxiliary contractions.

---

[50]Experimental psycholinguistic studies have found phrase frequency effects on production
(e.g. Bannard & Matthews 2008, Janssen & Barber 2012, Arnon & Cohen Priva 2013, Shao
et al. 2019); see Jacobs *et al.* 2016 for a review of frequency effects of word sequences in
multiple tasks.

First, special pronunciations appear only with the most frequent subjects. Bybee & Scheibman (1999:580) observe that in their *don't* data, though flapping of [d] occurs only with pronoun subjects, the further reduction of the [o] to [ə], occurs only with the subject *I*, the most frequent of the pronouns. Likewise, Table 3 illustrates pronunciations of tensed auxiliary contractions specific to the most frequent pronoun subjects, such as *I'll* [ɑl].

Second, Bybee & Scheibman (1999:590) observe that an adverb intervening between the subject and *don't* blocks vowel reduction (though it is not blocked by an adverb between *don't* and the verb). Likewise, the most reduced pronunciations of the subjects of restrictive auxiliary contractions are blocked by an intervening adverb:

(43) a.   *I'll* [aɪl/ɑl] *certainly come.*
           I [aɪ/*ɑ] *certainly 'll* [əl/*l] *come.*

      b.   *They're* [ðeɪɹ/ðɛɹ] *certainly expensive.*
           They [ðeɪ/*ðɛ] *certainly 're* [əɹ]/*ɹ] *expensive.*

Third, *don't* reduction fails with a conjoined pronoun *I* and with a lexical subject (Kaisse 1985, Scheibman 2000), as (44)a,b illustrate. (Following Scheibman (2000), the orthographic representation of reduced *I don't know* as *I dunno* is used here.)

(44) a.   **John and I dunno.*

      b.   **Those people dunno.*

The same syntactic restrictions characterize the restrictive contractions, as already seen in examples (23a–c).

The illustrative lexical entries in (45)–(47) are sufficient to capture all three properties of parallelism between contraction and *I dunno* reduction:[51] (1) the dependence on the specific pronoun *I* for the pronunciation of *don't* as [ɾə̃], (2) the required adjacency of *I* and *don't* for this reduced pronunciation, and (3) the syntactic restrictions against a conjoined subject with *I*, (44)a, and against a lexical noun phrase subject, (44)b.

---

[51]Zwicky & Pullum (1983) provide evidence that *n't* is an inflectional affix; see also Huddleston & Pullum (2002).

(45) *don't* [dõt/dõ] ←   I

$$(\downarrow \text{TENSE}) = \text{PRES}$$
$$(\downarrow \text{POLARITY}) = \text{NEG}$$
$$\neg(\downarrow \text{SUBJ PERS}) = 3$$
$$\downarrow = \Downarrow$$

(46) *I don't* [aɪɾə̃] ←   D   I

$$(\downarrow \text{PRED}) = \text{'PRO'} \qquad (\downarrow \text{TENSE}) = \text{PRES}$$
$$(\downarrow \text{PERS}) = 1 \qquad (\downarrow \text{POLARITY}) = \text{NEG}$$
$$(\downarrow \text{NUM}) = \text{SG} \qquad \neg(\downarrow \text{SUBJ PERS}) = 3$$
$$\Downarrow = \downarrow \qquad (\downarrow \text{SUBJ}) =_c \Downarrow$$

(47) *I don't know* [aɪɾə̃noʊ] ←

D	I	V
$(\downarrow \text{PRED}) = \text{'PRO'}$	$(\downarrow \text{TENSE}) = \text{PRES}$	$(\downarrow \text{PRED}) = \text{'KNOW}\langle(\downarrow\text{SUBJ})\rangle\text{'}$
$(\downarrow \text{PERS}) = 1$	$(\downarrow \text{POLARITY}) = \text{NEG}$	$(\downarrow \text{SUBJ}) =_c \Downarrow$
$(\downarrow \text{NUM}) = \text{SG}$	$\neg(\downarrow \text{SUBJ PERS}) = 3$	
$\Downarrow = \downarrow$	$(\downarrow \text{SUBJ}) =_c \Downarrow$	

The lexical entry in (47) is visualized extensionally in Figure 13.
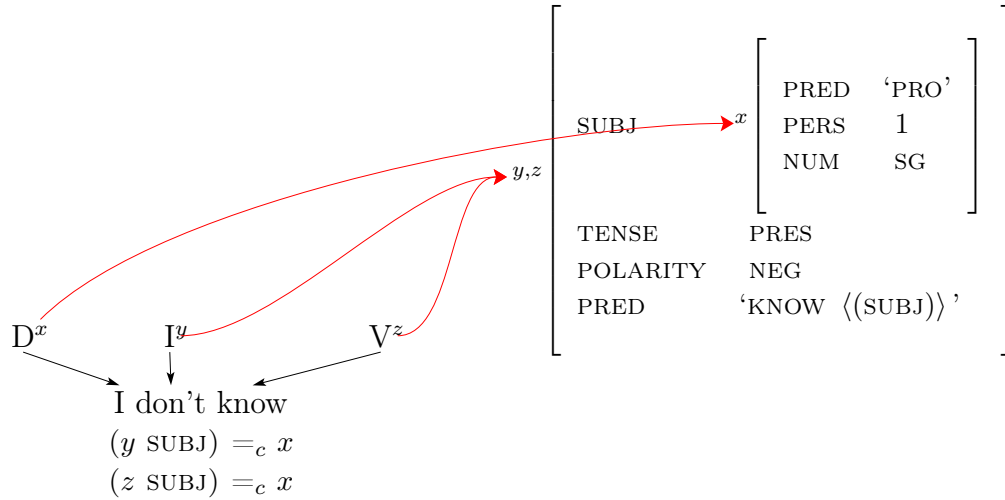

Figure 13: Visualization of the lexical entry for the unit *I don't know* (47).

Note that the verb *know* in (47) is specified intransitive, under the hypothesis that the special pragmatic functions associated with reduction require an

unspecified complement. The orthographic rendering *I dunno* seems to implicate this special pragmatic function. Compare (48)a,b, where the transitive use in (48)b seems less acceptable:

(48) a. *I dunno, Fred. I'm not sure I agree with you.*

b. *??I dunno Fred. Who is he?*

This intransitivity could be the reason for the reported ungrammaticality of examples (48)a,b, discussed by Scheibman (2000) and Kaisse (1985):

(49) a. **Tell me what you think I dunno ___ well enough.*

b. **The procedure that I dunno ___ involves applying to the grad school.*

The reduced instances of *I don't know* and the like are MULTI-WORD EXPRESSIONS. The analysis encapsulated in (46) and (47) shows that the theory of lexical sharing in principle allows the lexicalization of ANY STRINGS OF WORDS (collocations) which co-instantiate adjacent part of speech categories. This analysis extends LFG with lexical sharing from the quasi-morphological domain of portmanteau words and simple clitics squarely into the multi-word territory of usage-based linguistics.

The parallels outlined above suggest that what theoretically "triggers" lexical sharing in both constructions like tensed auxiliary contractions and multi-word expressions like *I don't know* is the same: the high usage probability of adjacent syntactic elements, just as Bybee and colleagues have argued. It is interesting that the lexical sharing of the small *I dunno* construction—which could be viewed at first glance as a grammatically isolated case—shows its usage-based character to be so similar to the lexical sharing of tensed auxiliary contractions, which are traditionally viewed as a systematic part of English grammar.

11   Concluding discussion

A central contribution of the present study is a high-level description of how a hybrid of formal grammar and the usage-based mental lexicon could explain the combined findings on tensed auxiliary contractions in English from both usage-based and formal lines of research. There are other architectures for

exemplar models of syntax that might be adopted. The dual-route multilevel exemplar model of Walsh *et al.* (2010) is noteworthy. Their key innovation is to explicitly formalize the relations between CONSTITUENTS and UNITS at both the phonetic and syntactic levels. For example, segments are constituents of syllable units and words are constituents of phrase or sentence units. These are stored in memory and categorized into into clouds of exemplars according to their similarity to existing exemplars. The architecture of their model employs two routes from every input to the output, setting up a competition between a submodel that directly selects the output as a unit exemplar and a submodel that assembles exemplar constituents into an output: the unit submodel wins if the unit exemplar receives activation above a threshold. Although discuss the goal of modeling phonetically detailed phrases stored in memory (e.g. Hay & Bresnan 2006) and in a related paper (Schütze *et al.* 2007) simulate the grammaticalization of *going to* ((Bybee 2006)), the Bybee-Pierrehumbert model adopted here more directly connects with the data of the present study.

Particularly interesting is that Walsh *et al.*'s (2010) models do not make any use at all of representational labels from formal grammar, whether phonological or syntactic. Their syntactic model achieves impressive results in learning grammaticality judgments of simple sentences (for example, *I like tea* vs. **I tea like*) from a purely quantitative distributional analysis of words in a corpus of child-directed speech to children of ages two to three years.[52] How this approach could extend to the complexities of adult grammatical knowledge remains to be seen. At bottom, all syntactic categories are distributional: "The similar syntactic behavior of two nouns like *coin* and *hen* is not directly apparent from their pronunciation or semantics. But an exemplar-theoretic account of syntactic behavior requires a similarity measure where *coin* and *hen* are similar" (Walsh *et al.* 2010:561–562). Although relational features like SUBJECT of course involve a much higher level of abstraction than sequential parts of speech (Bresnan *et al.* 2015), Walsh *et al.*'s (2010) multilevel exemplar model

---

[52]Building on a machine-learning approach to part-of-speech tagging (Schütze 1995), their model assigns each word two vectors, one consisting of the probabilities of all of their left-context words in the corpus and the other those of their right-context words, computed using relative frequencies that correspond to the maximum likelihood estimate for each probability. A word's similarity to exemplar words is measured by the sum of the cosines of these vectors (the same similarity measure used at the phonetic level in their syllable production model). In a simulation, Walsh *et al.* (2010) demonstrate that their distributional method of assigning fine-grained and gradient parts of speech to words performs better than category-based representations in judging the grammaticality of word order permutations of simple sentences.

is not fundamentally incompatible with the hybrid model sketched here.[53]

One other computationally explicit syntactic exemplar model is Bod's (1998, 2006, 2009) data-oriented parsing (DOP) model, in which the corpus IS the grammar. Bod & Kaplan (1998) and Bod (2006) show how the DOP model employing LFG c-structure to f-structure mappings can achieve productivity by parsing unseen data through structural analogy. In a very interesting later article, Bod (2009) shows how an unsupervised parser of data from the Eve corpus (Brown 1973) in the CHILDES database (MacWhinney 2000) can learn auxiliary inversion (a paradigm example of the seeming need for innate syntactic categories to overcome the "Poverty of the Stimulus" in language learning). The Pierrehumbert (2001) model adopted here provides a shorter and clearer path from the morphophonological data to the syntax of tensed auxiliary contraction.

The present study also makes an empirical contribution specific to the theory of LFG as a formal grammar in demonstrating the explanatory value of multiword lexically shared expressions (as does Broadwell 2007, n. 3).[54] Construction grammar (Fillmore *et al.* 1988; Goldberg 1995, 2006; Croft 2001) already allows lexical representations of multiword expressions as constructions, as does the data-oriented parsing (DOP) model. The formal grammar of the present study shares a number of linguistic features with Construction grammar, including the storage of lexically specific constructions (for example, Figures 4 and 13) and lexical schemata for productive constructions (30). Where Construction grammar aims to derive semantic distinctions among lexical words from their constructional contexts rather than from multiple lexical entries, the present study focuses on the usage-based lexicalization of syntactic fragments. There's no reason why the present framework could not be extended to other areas of grammar where usage affects the semantics and pragmatics of multiword expressions.

---

[53]In their conclusion Walsh *et al.* (2010:575) suggest that their model could be the basis for hybrid models of later stages of language development, with exemplar clouds linked to more abstract layers of representation, referring explicitly to the informal hybrid model of acquisition of Abbot-Smith & Tomasello (2006).

[54]Lowe (2016) proposes a dual-source analysis of the English *'s* genitive similar in spirit to the analysis of tensed *'s* of the present study: he assumes genitive *'s* is a clitic except in cases where lexical sharing with the host is motivated. However, his version of lexical sharing differs. Wescoat's λ mapping from c-structure to l(exical)-structure is a homomorphism which preserves linear order but not dominance, and supports a substantial version of the lexical integrity principle (Wescoat 2005, 2009). Lowe's π mapping is an inverse of λ and hence is a relation, not a function; it requires separately stipulating the linear order of atomic components of his shared entries, as well as the lexical integrity principle.

Bybee's conception of constructions in several works appears to eschew constituent structure. Bybee & Scheibman (1999) discuss the erosion of internal constituent structure boundaries associated with the phonetic fusions of frequently cooccurring words. While this erosion demonstrably occurs with frozen contractions in expressions like *whosie-whatsit, howsit/howzit* (n. 20), the evidence in Sections 4–5 shows that restrictive auxiliaries retain their constituent structure despite lexically specific phonetic fusions with their hosts. These contractions are intermediate between frozen lexicalizations and full syntactic phrases: they show phonetic compression and fusion, but retain syntactic life.

Bybee (2002) goes further to argue against hierarchical constituent structure altogether, proposing (p. 130), "Constituents of the type proposed for generative grammar which are described by phrase structure trees do not exist. Instead, units of language (words or morphemes) are combined into chunks as a result of frequent repetition." Her argument is based on the evidence that contractions like *you're* and similar units are chunks which overlap with c-structure constituents like NP and VP rather than nest hierarchically within them. In her view they consequently undermine the concept of hierarchical c-structure trees. However, the present study shows that *you're* can be both a lexical-syntactic unit or "chunk" and share a common c-structure with *you are*. The same is true of other common fragments such as *in the middle of* (Tremblay & Baayen 2010); see Bresnan (2001) or Bresnan *et al.* (2015) on the fragmentability of language in the LFG formal architecture.

The main contribution of the present study has been novel evidence for a hybrid formal and usage-based model of tensed auxiliary contractions. The novel evidence includes (i) a synthesis of the combined findings of formal and usage-based research on tensed auxiliary contraction, including their prosodic and metrical phonology, morphophonology, and syntax, and the relation of their usage probabilities to the likelihood of contraction, (ii) a corpus study of *is* contraction designed to test a crucial prediction of a hybrid formal and usage-based model, (iii) a formal analysis of the grammaticalization of host-auxiliary restrictions from their distributional usage patterns, and (iv) the extension of the formal grammar of auxiliary contraction to a multiword expression of classic usage-based grammar (Bybee & Scheibman 1999) that brings out surprising parallels with tensed auxiliary contraction. These results show the empirical and theoretical value of combining formal and usage-based data and methods into a more explanatory shared framework.

References

ABBOT-SMITH, KIRSTEN, & MICHAEL TOMASELLO. 2006. Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review* 23.275–290.

ABERCROMBIE, DAVID. 1961. Syllable quantity and enclitics in English. In *In Honor of Daniel Jones. Papers Contributed on the Occasion of his Eightieth Birthday, 12 September 1961*, ed. by David Abercrombie, D.B. Fry, P.A.D. MacCarthy, N.C. Scott, & J.L.M. Trim, 216–222. London: Longmans, Green, and Company, Ltd.

ALSINA, ALEX. 2010. The Catalan definite article as lexical sharing. In *Proceedings of the LFG10 Conference*, 5–25. Online: http://web.stanford.edu/group/cslipublications/cslipublications/LFG/.

AMBRIDGE, BEN. 2019. Against stored abstractions: A radical exemplar model of language acquisition. *First Language* 1–51. published by Sage Journals, DOI: 10.1177/0142723719869731.

ANDERSON, STEPHEN R. 2005. *Aspects of the Theory of Clitics*. Oxford: Oxford University Press.

——. 2008. English reduced auxiliaries really are simple clitics. *Lingue e linguaggio* 7.169–186.

ANTTILA, ARTO. 2017. Stress, phrasing, and auxiliary contraction in English. In *The Morpho-Syntax Phonology Connection*, ed. by Vera Gribanova & Stephanie Shih, 143–170. Oxford University Press.

——, TIMOTHY DOZAT, DANIEL GALBRAITH, & NAOMI SHAPIRO. To appear. Sentence stress in presidential speeches. In *Prosody in Syntactic Encoding. Special issue of Linguistisiche Arbeiten*, ed. by Gerrit Kentner & Joost Kremers. De Gruyter.

ARNON, INBAL, & URIEL COHEN PRIVA. 2013. More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and speech* 56.349–371.

BAKER, CARL L. 1971. Stress level and auxiliary behavior in English. *Linguistic Inquiry* 2.167–181.

BANNARD, COLIN, & DANIELLE MATTHEWS. 2008. Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological science* 19.241–248.

BARRON, JULIA. 1998. *Have* contraction: Explaining "trace effects" in a theory without movement. *Linguistics* 36.223–251.

BARTH, DANIELLE. 2011. Modeling reduction of *is*, *am* and *are* in grammaticalized constructions. *Quantitative Issues in Theoretical Linguistics (QITL)* 4.13.

——. 2019. Effects of average and specific context probability on reduction of function words *BE* and *HAVE*. *Linguistics Vanguard* 5. Published online by DeGruyter, 2019-05-08 DOI: https://doi.org/10.1515.

——, & VSEVOLOD KAPATSINSKI. 2017. A multimodel inference approach to categorical variant choice: Construction, priming and frequency effects on the choice between full and contracted forms of *am, are* and *is*. *Corpus Linguistics and Linguistic Theory* 13.203–260.

BENDER, EMILY M., & IVAN A. SAG. 2001. Incorporating contracted auxiliaries in English. In *Grammatical Interfaces in HPSG*, ed. by Ronnie Cann, Claire Grover, & Philip Miller, 1–15. CSLI Publications.

BOD, RENS. 1998. *Beyond Grammar: An Experience-based Theory of language*. CSLI.

——. 2006. Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review* 23.291–320.

——. 2009. From exemplar to grammar: A probabilistic analogy-based model of language learning. *Cognitive Science* 33.752–793.

——, & RONALD KAPLAN. 1998. A probabilistic corpus-driven model for lexical-functional analysis. In *Proceedings of the 17th international conference on Computational linguistics-Volume 1*, 145–151. Association for Computational Linguistics.

BÖRJARS, KERSTI, RACHEL NORDLINGER, & LOUISA SADLER. 2019. *Lexical-Functional Grammar. An Introduction*. Cambridge, UK: Cambridge University Press.

BRESNAN, JOAN, 1971. Contraction and the Transformational Cycle in English. Unpublished paper reproduced in 1978 by the Indiana University Linguistics Club.

——. 1973. Syntax of the comparative clause construction in English. *Linguistic Inquiry* 4.275–343.

——. 2001. *Lexical-Functional Syntax*. Oxford: Blackwell.

——. 2016. Linguistics: The garden and the bush. *Computational Linguistics* 42.599–617.

——, ASH ASUDEH, IDA TOIVONEN, & STEPHEN WECHSLER. 2015. *Lexical-Functional Syntax, 2nd Edition*. Blackwell Texbooks in Linguistics. John Wiley & Sons.

——, ANNA CUENI, TATIANA NIKITINA, & R. HARALD BAAYEN. 2007. Predicting the dative alternation. In *Cognitive Foundations of Interpretation*, ed. by Gerlof Bouma, Irene Krämer, & Joost Zwarts, 69–94. Amsterdam: Royal Netherlands Academy of Science.

——, & JESSICA SPENCER, 2012. Frequency effects in spoken syntax: *Have* and *be* contraction. Paper presented at the symposium on New Ways of Analyzing Syntactic Variation, Radboud University Nijmegen, 15–17 November 2012.

BROADWELL, GEORGE AARON. 2007. Lexical sharing and non-projecting words: The syntax of Zapotec adjectives. In *Proceedings of the LFG07 Conference*, 87–106. Online: http://web.stanford.edu/group/cslipublications/cslipublications/LFG/.

——. 2008. Turkish suspended affixation is lexical sharing. In *Proceedings of the LFG08 Conference*, 198–213. Online: http://web.stanford.edu/group/cslipublications/cslipublications/LFG/.

BROWN, ROGER. 1973. *A First Language: The Early Stages*. Harvard University Press.

BRYSBAERT, MARC, MICHAËL STEVENS, PAWEL MANDERA, & EMMANUEL KEULEERS. 2016. The impact of word prevalence on lexical decision times: Evidence from the Dutch lexicon project 2. *Journal of Experimental Pshychology: Human Perception and Performance* 42.441–458.

Bybee, Joan. 1985. *Morphology: A Study of the Relation between Meaning and Form*. Amsterdam: Benjamins.

——. 2000. The phonology of the lexicon: Evidence from lexical diffusion. In *Usage-based models of language*, ed. by Michael. Barlow & Suzanne Kemmer, 65–85. Stanford: CSLI.

——. 2001. *Phonology and Language Use*. Cambridge: Cambridge University Press.

——. 2002. Sequentiality as the basis of constituent structure. In *The Evolution of Language from Pre-Language*, ed. by T. Givón & Bertram F. Malle, 109–132. Amsterdam: John Benjamins.

——. 2006. From usage to grammar: The mind's response to repetition. *Language* 82.711–733.

——. 2010. *Language, Usage, and Cognition*. Cambridge: Cambridge University Press.

——, & Paul J. Hopper (eds.) 2001. *Frequency and the Emergence of Linguistic Structure*. John Benjamins Publishing.

——, & James L. McClelland. 2005. Alternatives to the combinatorial paradigm of linguistic theory based on domain general principles of human cognition. *The Linguistic Review* 22.381–410.

——, & Joanne Scheibman. 1999. The effect of usage on degrees of constituency: The reduction of *don't* in English. *Linguistics* 37.575–596.

Croft, William. 2001. *Radical Construction Grammar: Syntactic Theory in Typological Perspective*. Oxford University Press on Demand.

Dalrymple, Mary, John J. Lowe, & Louise Mycock. 2019. *The Oxford Reference Guide to Lexical Functional Grammar*. Oxford: Oxford University Press.

Davies, Mark, 2008–. *The Corpus of Contemporary American English: 400+ million words, 1990–present*. Available on-line at http://www.americancorpus.org.

DIESSEL, HOLGER. 2007. Frequency effects in language acquisition, language use, and diachronic change. *New Ideas in Psychology* 25.108–127.

DIXON, ROBERT M. W. 1977. Semantic neutralization for phonological reasons. *Linguistic Inquiry* 8.599–602.

ERNESTUS, MIRJAM. 2014. Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua* 142.27–41.

FILLMORE, CHARLES J, PAUL KAY, & MARY CATHERINE O'CONNOR. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 501–538.

FRANK, AUSTIN, & T. FLORIAN JAEGER. 2008. Speaking rationally: Uniform information density as an optimal strategy for language production. In *The 30th Annual Meeting of the Cognitive Science Society (CogSci08)*, 939–944, Washington, D.C.

GELMAN, ANDREW, & YU-SUNG SU, 2018. *arm: Data Analysis Using Regression and Multilevel/Hierarchical Models*. R package version 1.10-1.

GERMAN, JAMES, JANET BRECKENRIDGE PIERREHUMBERT, & STEFAN KAUFMANN. 2006. Evidence for phonological constraints on nuclear accent placement. *Language* 82.151–168.

GODFREY, JOHN J., & EDWARD HOLLIMAN. 1997. *Switchboard-1 Release 2*. Philadelphia: Linguistic Data Consortium.

GOLDBERG, ADELE E. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. University of Chicago Press.

——. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford University Press on Demand.

GORDON, ELIZABETH, LYLE CAMPBELL, JENNIFER HAY, MARGARET MACLAGAN, ANDREA SUDBURY, & PETER TRUDGILL. 2004. *New Zealand English: Its Origins and Evolution*. Cambridge University Press.

HARMON, ZARA, & VSEVOLOD KAPATSINSKI. 2017. Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology* 98.22–44.

HARRELL JR, FRANK E. 2001. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis.* Springer.

——, 2018. *rms: Regression Modeling Strategies.* R package version 5.1-2.

HAY, JENNIFER, & JOAN BRESNAN. 2006. Spoken syntax: The phonetics of *giving a hand* in New Zealand English. *The Linguistic Review* 23.321–349.

HUDDLESTON, RODNEY, & GEOFFREY K. PULLUM. 2002. *The Cambridge Grammar of the English Language.* Cambridge University Press.

INKELAS, SHARON. 1991. *Prosodic Constituency in the Lexicon.* Garland Publications.

——, & DRAGA ZEC. 1993. Auxiliary reduction without empty categories: A prosodic account. *Working Papers of the Cornell Phonetics Laboratory* 8.205–253.

ITO, JUNKO, & ARMIN MESTER. 2018. Matching light elements. In *A Reasonable Way to Proceed. Essays in Honor of Jim McCloskey*, ed. by Jason Merchant, Line Mikkelsen, Deniz Rudin, & Kelsey Sasaki, 168–190. Santa Cruz, Berkeley, Chicago: https://escholarship.org/uc/item/7z29n70x.

JACOBS, CASSANDRA L, GARY S DELL, AARON S BENJAMIN, & COLIN BANNARD. 2016. Part and whole linguistic experience affect recognition memory for multiword sequences. *Journal of Memory and Language* 87.38–58.

JANSSEN, NIELS, & HORACIO A BARBER. 2012. Phrase frequency effects in language production. *PloS One* 7. Published online: doi: 10.1371/journal.pone.0033202.

KAISSE, ELLEN M. 1979. Auxiliary reduction and the derivation of pseudo-clefts. *Linguistic Inquiry* 707–710.

——. 1983. The syntax of auxiliary reduction in English. *Language* 59.93–122.

——. 1985. *Connected Speech: The Interaction of Syntax and Phonology.* New York: Academic Press.

Keller, Frank, & Mirella Lapata. 2003. Using the Web to obtain frequencies for unseen bigrams. *Computational linguistics* 29.459–484.

Kiesling, Scott, Laura Dilley, & William D Raymond. 2006. *The Variation in Conversation (ViC) Project: Creation of the Buckeye Corpus of Conversational Speech*. Dept. of Psychology, Ohio State University, available at http://www.buckeyecorpus.osu.edu.

King, Harold V. 1970. On blocking the rules for contraction in English. *Linguistic Inquiry* 1.134–136.

Klavans, Judith L. 1985. The independence of syntax and phonology in cliticization. *Language* 95–120.

Krug, Manfred. 1998. String frequency: A cognitive motivating factor in coalescence, language processing, and linguistic change. *Journal of English Linguistics* 26.286–320.

Labov, William. 1969. Contraction, deletion, and inherent variability of the English copula. *Language* 45.715–762.

Levelt, Willem JM, Ardi Roelofs, & Antje S Meyer. 1999. A theory of lexical access in speech production. *Behavioral and Brain Sciences* 22.1–38.

Liberman, Mark, & Alan Prince. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8.249–336.

Lowe, John J. 2016. English possessive *'s*: Clitic *and* affix. *Natural Language & Linguistic Theory* 34.157–195.

MacKenzie, Laurel E. 2012. *Locating Variation above the Phonology*. University of Pennsylvania: PhD. dissertation.

——. 2013. Variation in English auxiliary realization: A new take on contraction. *Language Variation and Change* 25.17–41.

MacWhinney, Brian. 2000. *The CHILDES Project: The Database*, volume 2. Psychology Press.

McElhinny, Bonnie S. 1993. Copula and auxiliary contraction in the speech of White Americans. *American Speech* 68.371–399.

MEL'ČUK, IGOR. 1988. *Dependency Syntax: Theory and Practice*. Albany: SUNY Press.

NATHAN, GEOFFREY S. 1981. What's these facts about? *Linguistic Inquiry* 12.151–153.

PHILLIPS, BETTY S. 1984. Word frequency and the actuation of sound change. *Language* 320–342.

PIANTADOSI, STEVEN T., HARRY TILY, & EDWARD GIBSON. 2011. Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences* 108.3526–3529.

PIERREHUMBERT, JANET B. 2001. Exemplar dynamics: Word frequency, lenition and contrast. In *Frequency Effects and the Emergence of Lexical Structure*, ed. by Joan Bybee & Paul Hopper, 137–157. Amsterdam: John Benjamins.

——. 2002. Word-specific phonetics. *Laboratory Phonology* 7.101–139.

——. 2006. The next toolkit. *Journal of Phonetics* 34.516–530.

PITT, MARK A., KEITH JOHNSON, ELIZABETH HUME, SCOTT KIESLING, & WILLIAM RAYMOND. 2005. The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication* 45.89–95.

R CORE TEAM, 2019. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

RICKFORD, JOHN R., ARNETHA BALL, RENEE BLAKE, RAINA JACKSON, & NOMI MARTIN. 1991. Rappin on the copula coffin: Theoretical and methodological issues in the analysis of copula variation in African-American Vernacular English. *Language Variation and Change* 3.103–132.

SADLER, LOUISA. 1998. English auxiliaries as tense inflections. *Essex Research Reports in Linguistics* 24.1–15.

SCHEIBMAN, JOANNE. 2000. *I dunno:* A usage-based account of the phonological reduction of *don't* in American English conversation. *Journal of Pragmatics* 32.105–124.

SCHÜTZE, HINRICH. 1995. Distributional part-of-speech tagging. In *EACL 7*, ed. by S. Aloney & E. W. Hinrichs, 141–148. Dublin: Morgan Kaufmann.

——, MICHEAL WALSH, BERND MÖBIUS, & TRAVIS WADE. 2007. Towards a unified exemplar-theoretic model of phonetic and syntactic phenomena. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 29.

SELKIRK, ELISABETH O. 1972. *The Phrasal Phonology of English and French*. MIT: PhD dissertation.

——. 1984. *Phonology and Syntax*. MIT Press.

——. 1996. The prosodic structure of function words. In *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*, ed. by J. Morgan & K. Demuth, 187–214. Lawrence Erlbaum.

——. 2011. The syntax-phonology interface. In *The Handbook of Phonological Theory*, ed. by John Goldsmith, Jason Riggle, & Alan Yu, 435–483. Oxford: Blackwell Publishing.

SELLS, PETER. 1993. Juncture and the phonology of auxiliary reduction in English. *University of Massachusetts Occasional Papers in Linguistics* 8.76–105.

SEYFARTH, SCOTT. 2014. Word informativity influences acoustic duration: Effects of contextual predictability on lexical representation. *Cognition* 133.140–155.

SHANNON, C.E. 1948. *A Mathematical Theory of Communication*. Bell System Technical Report.

SHAO, ZESHU, JEROEN VAN PARIDON, FENNA POLETIEK, & ANTJE S MEYER. 2019. Effects of phrase and word frequencies in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 45.147.

SIMONS, GARY F., & CHARLES D. FENNIG. 2018. *Ethnologue: Languages of the World, Twenty-first edition*. Dallas, Texas: SIL International. Online version: http://www.ethnologue.com.

SÓSKUTHY, MÁRTON, & JENNIFER HAY. 2017. Changing word usage predicts changing word durations in New Zealand English. *Cognition* 166.298–313.

SPARKS, RANDALL B. 1984. Here's a few more facts. *Linguistic Inquiry* 15.179–183.

SPENCER, ANDREW. 1991. *Morphological Theory: An Introduction to Word Structure in Generative Grammar*. Oxford: Basil Blackwell.

SPENCER, JESSICA. 2014. *Stochastic Effects in the Grammar: Toward a Usage-Based Model of Copula Contraction*. Stanford University: PhD dissertation.

STERNBERG, SAUL, RONALD L. KNOLL, STEPHEN MONSELL, & CHARLES E. WRIGHT. 1988. Motor programs and hierarchical organization in the control of rapid speech. *Phonetica* 45.175–197.

——, STEPHEN MONSELL, RONALD L. KNOLL, & CHARLES E. WRIGHT. 1978. The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In *Information Processing in Motor Control and Learning*, ed. by George E. Stelmach, 117–152. New York: Academic Press.

SWEET, HENRY. 1890. *A Primer of Spoken English*. Oxford: Clarendon.

SZMRECSÁNYI, BENEDIKT. 2004. On operationalizing syntactic complexity. In *Le poids des mots. Proceedings of the 7th international conference on textual data statistical analysis. Louvain-la-Neuve*, volume 2, 1032–1039.

——. 2005. Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1.113–150.

TODD, SIMON. 2019. *The Listener in Language Change*. Stanford University: PhD dissertation.

——, JANET B. PIERREHUMBERT, & JENNIFER HAY. 2019. Word frequency effects in sound change as a consequence of perceptual asymmetries: An exemplar-based model. *Cognition* 185.1–20.

Tremblay, Antoine, & R Harald Baayen. 2010. Holistic processing of regular four-word sequences: A behavioral and ERP study of the effects of structure, frequency, and probability on immediate free recall. In *Perspectives on formulaic language: Acquisition and communication*, ed. by David Wood, 151–173. New York: Continuum.

Walsh, Michael, Bernd Möbius, Travis Wade, & Hinrich Schütze. 2010. Multilevel exemplar theory. *Cognitive Science* 34.537–582.

Wedekind, Jürgen, & Ronald M Kaplan. 2012. LFG generation by grammar specialization. *Computational Linguistics* 38.867–915.

Wescoat, Michael T. 2002. *On Lexical Sharing*. Stanford University: PhD. dissertation.

——. 2005. English nonsyllabic auxiliary contractions: An analysis in LFG with lexical sharing. In *Proceedings of the LFG05 Conference*, 468–486. Online: http://web.stanford.edu/group/cslipublications/cslipublications/LFG/.

——. 2007. Preposition-determiner contractions: An analysis in Optimality-Theoretic Lexical-Functional Grammar with lexical sharing. In *Proceedings of the LFG07 Conference*, 439–459. Online: http://web.stanford.edu/group/cslipublications/cslipublications/LFG/.

——. 2009. Udi person markers and lexical integrity. In *Proceedings of the LFG09 Conference*, 604–622. Online: http://web.stanford.edu/group/cslipublications/cslipublications/LFG/.

Wilder, Chris. 1997. English finite auxiliaries in syntax and phonology. In *Clitics, Pronouns, and Movement*, ed. by James R. Black & Virginia Motapanyane, 321–362. John Benjamins.

Wright, Saundra K., Jennifer Hay, & Tessa Bent. 2005. Ladies first? Phonology, frequency, and the naming conspiracy. *Linguistics* 43.531–561.

Zhan, Meilin, & Roger Levy. 2018. Comparing theories of speaker choice using a model of classifier production in Mandarin Chinese. In *Proceedings of the 17th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1997–2005.

——, & Roger P. Levy. 2019. Availability-based production predicts speakers' real-time choices of Mandarin classifiers. In *Proceedings of the 41st Annual Meeting of the Cognitive Science Society*.

Zwicky, Arnold M. 1970. Auxiliary reduction in English. *Linguistic Inquiry* 323–336.

——, 1977. On clitics. Unpublished paper reproduced in 1977 by the Indiana University Linguistics Club.

——, & Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English *n't*. *Language* 502–513.